Complex signal regulation drives the Arabidopsis immune network's response to bacterial flagellin stimulus

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Rachel A. Hillmer

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Fumi Katagiri, adviser

April 2016

Acknowledgements

My heartfelt gratitude to my Ph.D. adviser, Fumi Katagiri. Fumi is brilliant, kind, conscientious, and he helped me find my own feet as a scientist. When our ambitious modeling project took far longer than we expected (and I realized I wouldn't be the one to finish it), he wholeheartedly supported my suggestion of a new side project. Fumi, thank you for showing me by example what requiring excellence in all the details of one's work looks like. I am the statistician and the scientist I am today in no small part because of you. Thank you.

Thank you to Kenichi Tsuda for spending many hours training me in laboratory techniques. Your generous, open-hearted way of doing science inspires me.

Thank you to my committee members, who supported me through the long Ph.D. process. Thank you to Jane Glazebrook for your kindness and warmth. You showed me by example that a scientist can be completely down-to-earth and scientifically excellent all at the same time. Thank you to Chad Myers for challenging me to think more deeply about data interpretation. It was because of your encouragement that we discovered the tetra-stable switch motif in Yungil's paper, and my flg22 transcriptome paper is clearer and stronger because of you. Thank you for welcoming me into your own group meetings, and for traveling to the St. Paul campus many, many times to listen to my research progress and give me feedback. Thank you to Les Szabo for being a kind and steady support. I know you always had my back. And the stem rust microscopy images I got to make in your lab are my favorite. Best rotation project ever. Thank you to Igor Libourel for always being inquisitive and honest and sharp. I appreciate how carefully and rigorously you think about things.

Finally, thank you to Melody Olson and Sarah Hillmer. I made it through to the other side of this Ph.D. because of you two. When things were dark, you helped me find the light. When I was ready to give up, you carried me. You believed that this day would come long before I did. Thank you.

Abstract

Systems biology is the study of how biological systems operate as a whole. Systems become complex when interactions between system parts dominate system behavior. To uncover the mechanisms by which complex biological systems operate, those interactions must be discovered and quantified. Further, to understand dynamic system behavior, mechanistic rules for how system parts are stimulated and regulate each other must be discovered.

The plant immune signaling network, which protects plants from pathogens, is an especially complex system. Pathogens disable plant immune signaling with effectors; thus plant immunity must be robust against pathogen perturbation. Thus, deciphering the mechanisms that underlie the plant immune signaling network is met by a challenge: effects of single-gene mutations, on which traditional genetic analysis depends, are also buffered by the network.

In this dissertation, a network reconstitution approach was taken, where the network is disassembled and then stepwise re-assembled, to accurately assign network functions to system parts, including interactions between parts. We define the plant immune signaling network in terms of 4 major signaling sectors controlled by the plant hormones jasmonate (JA), ethylene (ET), and salicylate (SA) sectors, and the major immune regulator phytoalexin-deficient 4 (PAD4). Dynamic transcriptome and hormone profiles after plant immune stimulus with bacterial flagellin were collected across a combinatorially complete set of mutants, lacking all combinations of these four sectors. These mutant profiles were used in (1) attempts to find mechanistic mathematical models of immune network behavior and to (2) characterize the four-sector network's control of the flg22-responsive transcriptome.

The work in this dissertation produced two main discoveries. First, that delay differential equations (DDEs) can be found which provide mechanistic explanations of immune network function; additional time course detail will be needed to confirm the accuracy of these models. Second, network buffering is extensive in the flg22-responsive transcriptome. As a result of this network buffering, our network reconstitution based interpretations of gene regulation are at points quite different from the regulatory mechanisms described in the plant immunity literature.

Table of Contents

Acknowledgements i
Abstract ii
List of Tables vi
List of Figures vii
Chapter 1. Introduction: System Modeling of Dynamic Plant Processes 1
1.1 Mathematical modeling of plant signaling networks 3
1.2 Modeling the plant immune signaling network's dynamics 4
Chapter 2. Systems Biology for Biologists. 7
2.1 Have you been put off by systems biology? 7
2.2 What is a system? 7
2.3 What about some biological systems? 7
2.4 My system has tens of thousands of molecular parts. Aren't your claims that systems biology will help me just wild speculation? 8
2.5 How do I define the rules governing my biological system? 9
2.6 What is a mathematical model? 9
2.7 Are there any systems biology success stories? 10
2.8 Should you become a systems biologist? 11
2.9 Acknowledgements 11
Chapter 3. A search for mechanistic mathematical models of the JA-ET-PAD4-SA network 14
3.1 Chapter Overview 14
3.2 Background: Coarse-grained empirical modeling of the Arabidopsis plant immune signaling network 15
3.3 ND2D models of the JA-ET-PAD4-SA network 17
3.3.1 Selection of sector activity marker genes 18
3.3.2 Interpolation of sector activity data 18
3.3.3 ND2D model structure discovery 28
3.3.4 ND2D model structure and parameter selection for all double mutants 32
3.3.5 ND2D example rule composition: JA→SA and PAD4→SA. 38
3.3.6 Additive combinations of ND2D models 41
3.4 A differential-equation-based framework for modeling the JA-ET-PAD4-SA network. 43
3.4.1 Calculating concentration-based sector activity values and derivatives 43
3.4.2 Delay Differential Equations (DDEs) can describe the triple mutant sector activities 46

3.4.3 A DDE-based model of PAD4 sector activity 51
3.5 Future Directions 59
Chapter 4. The Highly Buffered Arabidopsis Immune Signaling Network Hides the True Function of its Components 62
4.1 Summary 63
4.2 Introduction 64
4.3 Results and Discussion 66
4.3.1 Overview of data collection and analysis 66
4.3.2 The vast majority of the flg22-responsive and network-dependent genes showed complex network regulation 67
4.3.3 About half of the flg22-responsive and network-dependent genes are well buffered. 71
4.3.4 The four sectors show contrasting patterns of network dependence 73
4.3.5 Network reconstitution revealed unexpected patterns of transcriptional and hormone regulation 78
4.3.6 Network reconstitution revealed complex regulatory mechanisms that were hidden from null-mutant analysis of the individual genes in the 4-sector network 80
4.3.7 Genes in the ERF-branch of JA signaling did not display the expected strong synergy between JA and ET 84
4.3.8 Promoter motif analysis of gene sets with similar signaling allocations 87
4.3.9 Conclusions 89
4.4 Materials and Methods 89
4.4.1 Plant materials, growth conditions, leaf tissue treatment and collection 89
4.4.2 Hormone measurement 90
4.4.3 Tag-Seq library preparation and measurement of gene transcript expression levels 91
4.4.4 Gene selection 91
4.4.5 Signaling allocation models 92
4.4.6 MEME-based motif discovery 93
4.4.7 Clustering 93
4.4.8 Gene Ontology (GO) enrichment analysis 93
4.4.9 Data submission to NCBI's Gene Expression Omnibus 93
4.5 Acknowledgements 93
4.6 Supporting Information 94
4.6.1 S1 Text. Supporting Methods. 94
4.6.2 S2 Text. Motif Analysis Summary. 107
4.6.3 Supporting Figures and Tables 109

4.6.4 Additional Supporting Information....132
Chapter 5. Conclusion....135
5.1 Major Findings....135
5.2 Future Directions....136
5.3 Concluding Remarks....137

List of Tables

Chapter 3

Table 3-1. Coarse-grained model search parameter values 35
Table 3-2. Parameters for the best fitting model for each sector's effect on every other sector's activity 38
Table 3-3. Best adjusted R^2 values for additive combinations of ND2D models in wild type and all single-gene mutants 42
Table 3-4. Parameter values for the PAD4 DDE model optimization 57

Chapter 4

S1 Table. Figure 4-6 cluster C and cluster D genes include genes with SA-related functions.....132

List of Figures

Chapter 2

Figure 2-1. Combinations of system perturbations are required to learn the mechanisms underlying complex biological system responses 12

Chapter 3

Figure 3-1. The time axis transformations used for polynomial interpolation 21
Figure 3-2. Pseudo time interpolations of JA sector activity data 22
Figure 3-3. Pseudo time interpolations of ET sector activity data.....23
Figure 3-4. Pseudo time interpolations of PAD4 sector activity data 24
Figure 3-5. Pseudo time interpolations of SA sector activity data.....25
Figure 3-6. Time-interpolated JA sector activity values 26
Figure 3-7. Time-interpolated ET sector activity values.....27
Figure 3-8. Time-interpolated PAD4 sector activity time courses 27
Figure 3-9. Time-interpolated SA sector activity time courses.....28
Figure 3-10. A very well-behaved input rule was found for the PAD4 effect on the SA sector in the double mutant *jePS* 32
Figure 3-11. Diagrams of the family of ND2D models that I considered when searching for rules to explain double-mutant sector activity behavior.....33
Figure 3-12. Best-fitting ND2D model for each 2-sector relationship, fit to time-interpolated double mutant and triple mutant sector activity data.....36
Figure 3-13. A single PAD4→SA model structure and parameter set can explain SA sector responses in both *jePS* and *JePS* relative to SA in *jepS*.....39
Figure 3-14. Performance of *JePS* data using the best fitting PAD4→SA *jePS* model 40
Figure 3-15. A single PAD4→SA rule can explain SA sector responses in all genotypes where PAD4 and SA are active 41

Figure 3-16. Time-interpolated JA and ET sector activity data and their derivatives, in concentration units (y-axis) 44
Figure 3-17. Time-interpolated PAD4 and SA sector activity data and their derivatives, in concentration units (y-axis) 45
Figure 3-18. A DDE model fits time-interpolated JA activity values in the triple mutant *Jeps* very well.....49
Figure 3-19. A DDE model fits time-interpolated ET activity values in the triple mutant *jEps* very well. 50
Figure 3-20. A DDE model fits time-interpolated PAD4 activity values in the triple mutant *jePs* very well 50
Figure 3-21. A diagram of our DDE-based model of PAD4 sector activity 52
Figure 3-22. A DDE model fits time-interpolated PAD4 activity values in all PAD4-containing genotypes well 56

Chapter 4

Figure 4-1. Network sector interactions are common in the regulation of the transcriptome response to flg22 treatment.....69
Figure 4-2. Network sector interactions dominate over individual single sector contributions to the flg22 transcriptome response.....70
Figure 4-3. The responses of many genes to flg22 treatment are highly buffered by the network.....72
Figure 4-4. SA hormone induction completely depends on the other sectors 74
Figure 4-5. No genes are regulated by the SA sector alone.....75
Figure 4-6. Two clusters of genes under distinct regulatory mechanisms are within the traditionally-defined SA-dependent genes 77
Figure 4-7. Regulation of genes strongly impacted by the JA and ET single sector perturbations 82
Figure 4-8. ERF-branch genes do not show the expected dominance of the J:E interaction.....86

Figure 4-9. Several genes show dominance of the J:E interaction 88
S1 Fig. Signaling allocation models for Figure 4-1 genes and hormones without q-value-based regularization 109
S2 Fig. Fraction of interaction-based contributions to the signaling allocations of genes and hormones in S1 Fig 110
S3 Fig. Transcriptional behavior of flg22-responsive genes in *fls2*, the wild type, and the quadruple mutant 111
S4 Fig. Signaling allocation for an alternative ET sector activity marker, *ARL* (AT2G44080).....112
S5 Fig. Signaling allocations for alternative PAD4 sector activity markers 113
S6 Fig. A time course of the $\log_2$ free SA hormone concentration per genotype 114
S7 Fig. A time course of the adjusted $\log_2$ free JA hormone concentrations per genotype 115
S8 Fig. *FLS2* transcript levels in the *ein2*-containing genotypes.....116
S9 Fig. Figure 4-6 without q-value based regularization, and without masking insignificant changes in *sid2* 117
S10 Fig. Model correlations with *sid2* transcript response changes relative to wild type for traditionally-defined SA-dependent genes 118
S11 Fig. Few genes show differential expression between wild type and *pad4* at any sampled time point 119
S12 Fig. An early ethylene spike is evident in ET sector activity marker genes 120
S13 Fig. *PR4* was down-regulated at early time points after flg22 treatment 121
S14 Fig. Transcript expression level responses for the genes and hormones in Figure 4-1.....122
S15 Fig. Transcript expression dynamics for most genes are very similar across the combinatorial sector genotypes 123
S16 Fig. Genotype differences in transcript expression dynamics are evident in flg22-induced transcript response changes 124

S17 Fig. Genotype differences in transcript expression dynamics are evident in transcript response changes.....125
S18 Fig. Time courses of free SA and free JA hormone concentrations per genotype 126
S19 Fig. Representative GA-rich and AAG-rich motifs found in the promoters of flg22-responsive, network-responsive genes.....127
S20 Fig. Histogram of sample scaling factors for all 357 Tag-Seq datasets 127
S21 Fig. Mean residual variance for the searched range of a parameters 128
S22 Fig. Residual variance for mean-ranked Tag-Seq data.....129
S23 Fig. Calculation of transcript response change values for the gene PR4 130
S24 Fig. Correlation between model-predicted data and actual data.....131
S25 Fig. Correlation between model-predicted data and actual data.....131

Chapter 1.
Introduction: System Modeling of Dynamic Plant Processes

"A system is more than the sum of its parts." –Aristotle.

When multiple factors and components contribute to a biological process of interest, it is often the case that these system parts interact with each other. In the case that these interactions are strong, the system that is governed by these interacting parts becomes complex; the system is no longer merely the sum of independent part functions. To understand a complex system, the interactions between its parts must be discovered and quantified. This is the goal of systems biology. The aim of the work described in this dissertation was to discover and quantify salient interactions in the plant immune signaling network.

The response of a biological network to a signal is inherently dynamic, since biological and environmental signals are themselves dynamic. Thus, good models of the biological systems that process information must be dynamic. Understanding how interactions drive system dynamics is a key to building system-level understanding of biological networks and the processes they control. Thus a major feature of this dissertation is the analysis of time course data after immune stimulus.

Plant systems biology is a relatively underdeveloped field. Plants grow slowly compared to *Escherichia coli* and *Saccharomyces cerevisiae*, the model systems in which the lion's share of systems biology studies have been performed. Plant model systems are also largely less developed than animal model systems; fewer molecular details of important signal transduction pathways have been identified in plants than in animal systems. Nevertheless, research in plant systems biology is vital: agricultural improvements and food security are on the line. Pathogen drag is a major component of agricultural yield loss (Oerke 2005; Strange and Scott 2005; Bebber, Ramotowski, and Gurr 2013; Bebber and Gurr 2015). Understanding the mechanisms by which the plant immune signaling network operates may suggest new modes of plant protection. This work in this

dissertation contributes to our understanding of the mechanisms that underlie the plant immune signaling network; this network protects plants from pathogens.

This dissertation is also a contribution to the systems biology literature. The networks that control many biological processes are not only complex but also large, consisting of at least hundreds of different components (Fischer 2008). Most of the successful mechanistic mathematical models of dynamic biological systems to date have used the equations of Michaelis-Menten (MM) kinetics to directly model molecular reactions. However, when the number of system components exceeds a dozen or so, this approach becomes intractable. There are several reasons why. First, as the system becomes large, the number of kinetic parameters which must be estimated from the data also becomes large; to estimate these parameters requires system perturbations that decouple these parameters from each other. Finding appropriate perturbations and collecting the data needed to support MM modeling for large biological networks is a serious hurdle. Second, fitting large systems of differential equations, testing for parametric sensitivity, and simulating the behavior of these equations under various initial conditions is very computationally intensive when more than a dozen equations are in play; an example of these computational difficulties is demonstrated in efforts to model the Drosophila segmentation network using 17 ordinary differential equations (von Dassow et al. 2000; von Dassow and Odell 2002). Furthermore, MM kinetics is a continuous approximation of a statistical process; including stochastic events in ODE models can be done, but with even greater computational challenges (Székely and Burrage 2014). The practical choice therefore arises between either modeling relatively small sub-networks (Alon 2006), or choosing a summarized scale at which to model the full network. The second involves discovering the system rules *de novo*. This second approach has been used successfully in ecology, epidemiology, and physiology (Ellner and Guckenheimer 2011), but is underutilized in molecular genetics. Nevertheless, molecular genetics is an ideal domain for empiricism; precise tools exist for removing, adding and modifying components of interest. This dissertation seeks to discover, *de novo*, the rules by which four summarized signaling sectors of the model plant *Arabidopsis thaliana* regulate each other and the full plant transcriptome in response to the immune stimulus bacterial flagellin. These four sectors are the jasmonate (JA), salicylate (SA), ethylene (ET), and phytoalexin-deficient 4 (PAD4). They mediate the

signaling from three major plant hormones (JA, ET, and SA), and from the major immune regulator PAD4.

1.1 Mathematical modeling of plant signaling networks

To set my dissertation work in context, I begin with a brief overview of the mechanistic mathematical models of plant signaling networks that have been published to date. There are few such models. Moreover, mechanistic mathematical models for plant signaling responses to biotic pressure are to date lacking from the literature; models have been made for animal innate immunity (Pigozzo et al. 2013), but not for plant immunity. Published mechanistic models of plant signaling networks are mostly based on MM kinetics; the scale of these models is bound to molecular events where the MM equations can be applied. This dissertation aims to help bridge a two-fold gap in the plant science literature: I sought for mechanistic mathematical rules of plant immune function, while also using a more summarized modeling scale.

Most of the highly developed plant mechanistic mathematical models are models of core plant processes. These include models of core photosynthesis and starch metabolism (Pettersson and Ryde-Pettersson 1988; Laisk, Eichelmann, and Oja 2006; Nägele et al. 2010), including non-equilibrium models (Poolman 2000). Metabolic models are sufficiently advanced that they can be used for directed engineering (McNeil 2000; Libourel and Shachar-Hill 2008). However, these metabolic models have an advantage over models of signaling networks: a strong set of constraints from conservation laws of stoichiometry (Maarleveld et al. 2013).

There are far too many unknowns in signal transduction networks to rely on this constraint-based approach. Nevertheless, several notable plant signaling networks have been modeled using MM kinetics. Dynamic mathematical models have been used to capture the mechanisms underlying the plant circadian clock (Alabadí et al. 2001). Clock models have benefitted from iterative rounds of modeling and experimentation, which identified missing loops (J C W Locke, Millar, and Turner 2005; James C W Locke et al. 2005; James C W Locke et al. 2006), and allowed investigation into emergent network features, like plasticity and controllability (Dodd et al. 2014; Dalchau et al. 2010).

Mathematical models have also been successfully used to model dynamic patterning in the Arabidopsis apical shoot meristem and to identify previously unknown feedback loops (Gordon et al. 2009; Chickarmane et al. 2012). Additionally, a mechanistic mathematical model has been used to describe the Arabidopsis floral transition (Jaeger et al. 2013). Finally, a dynamic mathematical model has been built for jasmonate (JA) biosynthesis feedback and the resulting transcriptional response that occurs after an elevation in JA levels (which often occurs during herbivory) (Banerjee and Bose 2011).

A major signal processing feature of many plant processes is that they must be able to buffer environmental noise, ensuring operation of critical plant functions despite the continually fluctuating environmental conditions to which most plants are exposed. Meanwhile, these networks must still respond to and integrate external cues, notably, light cues. Plant signaling networks which respond to potential pathogens face an even more complex task. Not only must they respond reliably to pathogen attack signals, while avoiding mis-firing, as inducible immunity is costly (Vos, Pieterse, and Wees 2013), but immune signals must be reliably transduced into immune responses while pathogens actively engage in disabling the plant processes that mediate immune signals, using secreted or injected effectors. Plants face a distinct disadvantage compared to their microbial pathogens: microbes evolve much faster than plants. Thus, the plant immune network must be robust against attack and resilient against pathogen evolution of new effectors (Tsuda and Katagiri 2010).

1.2 Modeling the plant immune signaling network's dynamics

There is detailed, and expanding, knowledge of how immune stimuli are perceived, that is, for how plant immune signal transduction is initiated (Dodds and Rathjen 2010; Jones and Dangl 2006; Dangl and Jones 2001). Effector-triggered immunity (ETI), which triggers a fast and strong form of immunity, relies on the direct or indirect recognition of pathogen effectors by plant Resistance (R) proteins. Pathogen effectors are highly diversified. On the other hand, molecular moieties that are highly conserved across broad classes of microorganisms also trigger plant immunity. Bacterial flagellin, bacterial elongation factor EF-Tu, and chitin (a component of fungal cell walls), among other

microbe-associated molecular patterns (MAMPs), trigger a broad-spectrum form of immunity called Pattern-Triggered Immunity (PTI). Receptors and co-receptors for several MAMPs have been identified (Newman et al. 2013). Several key immune signaling events downstream of pathogen perception are also well known, including MAP Kinase activation, reactive oxygen species (ROS) production, calcium signaling, and defense gene activation. Yet mechanistic models that can predict immune signaling dynamics and the phenotypic outcomes of this signaling with accuracy do not yet exist.

The work in this dissertation builds on previous genetic work that defined the very large and complex plant immune network in terms of four subnetworks: the salicylate (SA), jasmonate (JA), ethylene (ET), and phytoalexin-deficient 4 (PAD4) sectors (Tsuda et al. 2009). This four-sector network explains 80% of the flg22-PTI phenotype against the bacterial pathogen *Pseudomonas syringae* pv. *tomato* DC3000. The models in this previous study (Tsuda et al. 2009) were statistical models, with causal connections between genotype and phenotype. This work (Tsuda et al. 2009) was followed by a semi-dynamic model built from sector activity profiles at two time points after MAMP treatment (Kim et al. 2014). This semi-dynamic model revealed the signaling flows between sectors that drive network behavior after MAMP treatment. The work in this dissertation sought to reach the next milestone in mechanistic modeling, discovering rules for sector interactions that could predict the signal flows seen in the Kim model (Kim et al. 2014). The data which informed the modeling in this dissertation were detailed time course sector activity profiles after plant treatment with flg22, a potent MAMP; data were collected by collaborators. While promising delay-differential equation models were found (see Chapter 3), the complexity of these models approached, and in some cases exceeded, the complexity of the underlying data. These models were fit to time-interpolated data, thus, while the models explained the time-interpolations very well, they are likely fitting interpolation artifacts in addition to the network's underlying mechanisms. Additional time resolution where the underlying data is dynamic will be needed to verify that model terms accurately and only reflect network mechanisms, and not time-interpolation artifacts.

Detailed transcriptome profiles after flg22 treatment were collected by collaborators to extract the input data for the modeling work in Chapter 3. In Chapter 4, I studied the

dynamic mechanisms by which the JA-ET-PAD4-SA network regulates the genes in the flg22-responsive transcriptome. While Chapter 3 asks, “What are the mechanisms underlying JA-ET-PAD4-SA network behavior?”, Chapter 4 asks, “How does the control network, that is, the JA-ET-PAD4-SA network, regulate the transcriptome?”. Chapter 4 provides extensive documentation of just how much the plant signaling immune network really is more than the sum of its parts; I described and investigated the extensive network buffering that is a key consequence of the beyond-additive mechanisms that underlie the plant immune signaling network.

The work in this dissertation begins with an introduction to systems biology (Chapter 2), a primer I wrote on why a systems approach that can characterize both system parts—and, critically, their interactions—is vital in discovering the regulatory mechanisms of complex networks.

Chapter 2.
Systems Biology for Biologists.

 The Pearls are a collection of teaching articles for “lessons that last.”

2.1 Have you been put off by systems biology?

Do you avoid papers thick with mathematical details and unfamiliar statistical analyses? If so, this article is for you! Systems biology, at its core, is not a set of computational and mathematical techniques; these are merely tools, incredibly useful, but secondary. The heart of systems biology is simple: explaining how a system works requires an integrated outlook. For any phenotype—molecular, macroscopic or ecological—a set of interrelated factors exist that contribute to this phenotype. Since these factors interact, they need to be studied collectively, not merely individually. That’s it!

2.2 What is a system?

A system is a collection of parts and factors that work together to complete a task. Conversely, for a given task, the system is defined by the set of all parts and factors which influence, accomplish, or impede that task.

Some systems are easy to identify. Think of a machine, like a car. The body of the car houses all the parts that make up the automobile. The external boundary makes the system easy to identify. Some systems are less easily identified. Consider all the factors that influence traffic flow in a city. The first example is concrete, the second more abstract; both are systems.

2.3 What about some biological systems?

In the systems biology literature, the most commonly discussed systems are networks of genes or proteins. Sometimes these are very large systems: the set of all genes in an

organism (e.g. (Lim et al. 2005; Eichten et al. 2013; Costanzo et al. 2010)). But there is no fixed scale at which systems biology operates. Your system could be an ecosystem of plants, and the soil services they provide and require; your system could be an epidemiological system with hosts, pathogens, and vectors. Your system could be a single molecular process, like the regulation of an important gene. Or it could be complex system like induction of an immune response within a cell, tissue or organism. If there is a biological question you wish to ask, or a process you wish to study, there is, *de facto*, a set of parts which contribute to that process; these parts define the system. Biological parts are interconnected and interdependent. Systems biology recognizes this and provides tools and frameworks to both accurately capture these relationships and deduce the system behavior that emerges from these relationships.

2.4 My system has tens of thousands of molecular parts. Aren't your claims that systems biology will help me just wild speculation?

Happily, no. The solution: taking stock of major effects and ignoring minor ones. Good systems biology is a balance between reductionism—breaking a system apart into smaller parts and defining the function of these parts—and synthesis—understanding how the parts cooperate to produce the behavior of the whole. We have two options: (1) discover and study small modular sub-systems (Alon 2006) or (2) approximate a complex system via a tractable number of components (Ellner and Guckenheimer 2011) (e.g. (Kim et al. 2014)). To do the latter, we first look for the parts which have large effects on our process or phenotype of interest, so-called "large-effect" parts (Coe 2002). For example, if the system is an organism-level process, these may be hormone concentrations, which by definition regulate a large number of molecular processes, e.g.(Kim et al. 2014; Gordon et al. 2009). Or they may be cells in the circulatory system, which can monitor and regulate multiple tissues, e.g.(Kosmrlj et al. 2010). If we first get a good approximation of the basic functionality of a system, we can then add on the bells and whistles.

To figure out the major-effect parts and/or processes of your system, there are a host of established biological methods, including:

(1) observation

(2) forward genetic screens
(3) genome-sequence-assisted guesses
(4) co-expression analysis
(5) exogenous chemical application

2.5 How do I define the rules governing my biological system?

Once you have a first-pass parts list assembled, you will need to combine two types of experimental factors: (see e.g.(Ellner and Guckenheimer 2011; Kim et al. 2014))

(1) External system perturbations.
(2) Internal system perturbations.

Why are perturbations needed? As in classical genetics, we learn about systems best by breaking or aggravating them in defined ways, observing how those induced changes modulate the process or phenotype of interest. External perturbations include, for example, treating a tissue with pathogens or pathogen-derived compounds. An internal perturbation involves removing, disabling, or modifying one or more system parts. Diverse internal system perturbations are needed because complex, robust systems are often full of redundancies and backups. Robust systems buffer mild perturbations. Ideally, a combinatorial set of internal perturbations that jointly abolish a phenotype would be challenged by a representative diverse suite of external system perturbations that stimulate the system in different precise ways (Kim et al. 2014; Tsuda et al. 2009; Jansen 2003).

The system should be measured across appropriate timecourses to capture when the system is dynamically responding to the (esp. external) perturbations. Quantitative monitoring of both the system parts and the system output empowers mathematical deduction of the mechanisms by which system parts control and modulate the system response.

2.6 What is a mathematical model?

A mathematical model is a set of relationships, usually written as equations, that describe how the parts of system respond to system inputs, regulate each other, and

control system output. Why do we need math to do this? Math is just formalized logic, so in theory we could just use descriptive sentences. But for all but the simplest relationships, exhaustively working out the implications of these relationships is prohibitively laborious, and error-prone. Moreover, language can be imprecise, where math naturally tends towards precise expression of relationships. Why is a model valuable? Models are hypothesis-generation tools, efficient ways to scout out novel and interesting system behavior. We use them to explore *in silico* varied external conditions and internal system modifications. Accomplishing such exploration experimentally is usually far more labor intensive, costly, and perhaps even impossible. A good model is an imperfect but useful virtual copy of a system that reproduces the salient features of the system. This copy lets us play with the system using computational techniques, analogous to how physical toy models help chemists think about the structure of a molecule.

Perfect system knowedge is *not* a prerequisite for starting to build a mathematical model. Model building is an iterative procedure: model, predict, test experimentally, repeat all. Modeling, when done well, will help channel further experiments in the most fruitful directions.

Have no fear, you do not need to become an expert in math. You need only make friends with someone who is. And your collaborator very much needs your input on the model. Mathematical approaches and structures need to be chosen which capture and reflect the essence of each biological system. For this, the training and intuition of a biologist is irreplaceable.

2.7 Are there any systems biology success stories?

Why yes, indeed there are. The history of successful mathematical modeling in biology has a history much longer than that of the genomics era. Here are some highlights: Tissue models of the human heart stand on over half a century of iterative modeling, experimentation, and model refinement. Birthed from this long labor, the virtual heart effort, used in clinical settings, may be systems biology's brightest star (Freedman 2004; Kohl and Noble 2009). In 1952, British mathematician Alan Turing proposed that leopard

spots, zebra stripes, and spirals in nature did not require complicated signaling, but arise by a simple reaction-diffusion equation imposed on a homogenous system (Turing 1952). It took decades to develop the molecular tools to test his hypothesis, but he was right (Economou et al. 2012; Ouellette 2013). During the 2001 UK outbreak of foot-and-mouth disease, mathematical models were used to predict disease spread and assisted in deciding control measures (Keeling 2005). An integrated biomedical informatics program, aneurIST, predicts rupture of incidentally discovered cerebral aneurysms using additional patient-specific medical data. During active model development, it was estimated that this modeling effort saved thousands to millions of Euros annually in unnecessary procedures ("aneurIST" 2015). Nicolas Le Novère has an accessible description of several successful classical molecular systems biology models on his blog (Le Novère 2015). For numerous additional examples, see the European Bioinformatics Institute's (EMBL-EBI) 'model of the month' database, part of its BioModels database ("BioModels" 2015).

2.8 Should you become a systems biologist?

Are you perfectly content to study a small system? There's no pressure to take on a wildly ambitious system, understanding the function of an entire cell, or modeling the ecosystem of planet earth. For any biological question, a relevant system exists, and the study of this system will benefit from including mathematical models in your toolkit.
You may be hesitant to consider becoming a systems biologist. Math fear is a real thing. But who knows? There might be a collaborator waiting for you just across campus. From my experience as a physicist turned biologist, I can confidently say there are mathematicians, physical and computer scientists and engineers who are lured by the extraordinariness of biology.

2.9 Acknowledgements

I thank Dr. Heather L. True-Krob for encouraging me to write this article. I gratefully acknowledge Dr. Fumi Katagiri, my thesis adviser, for his influence on my perspective of systems biology. I thank the reviewers for their clear and constructive comments on this article. I also thank the University of Minnesota Theory Group

(http://eeg.github.io/theory_lab/) for their many excellent suggestions of success stories of ecological-scale mathematical models.

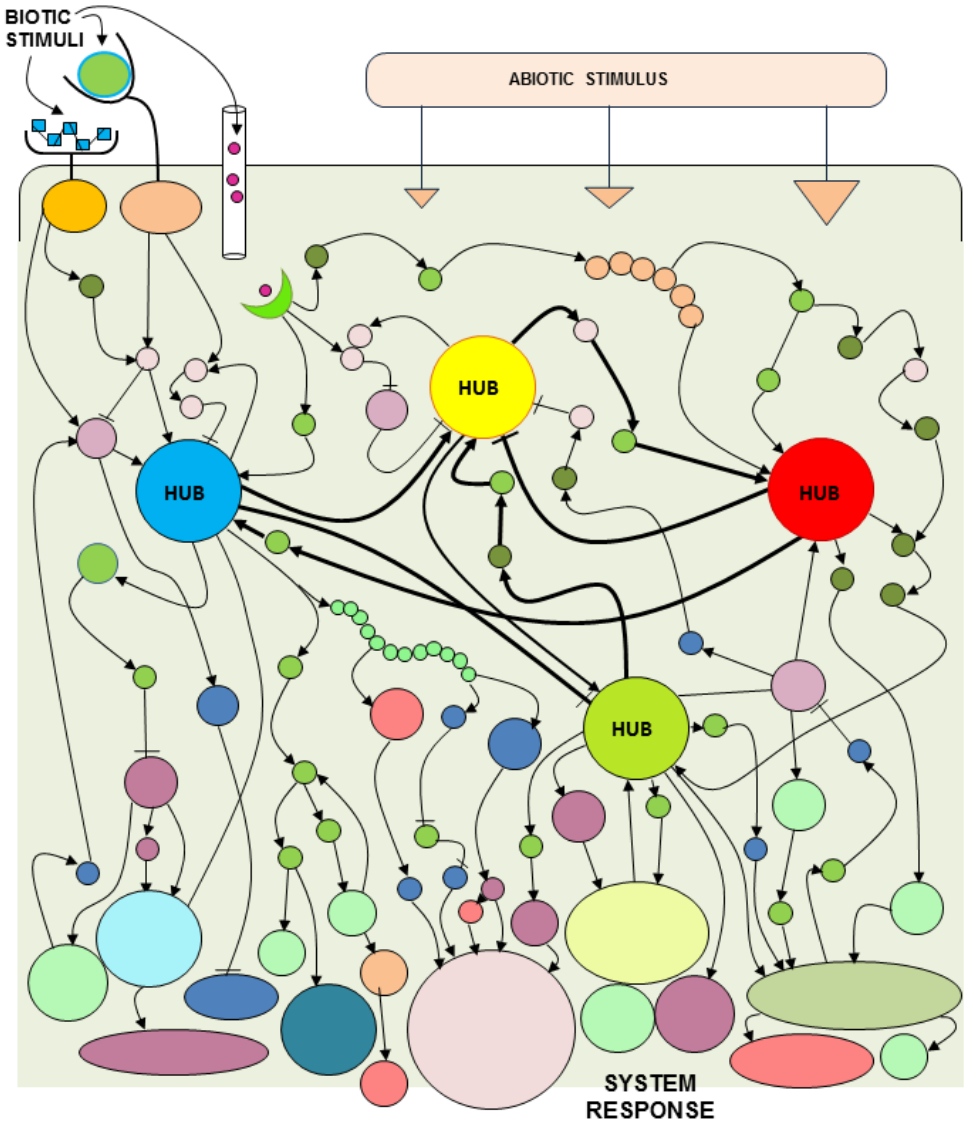


Figure 2-1. Combinations of system perturbations are required to learn the mechanisms underlying complex biological system responses. Complex systems, e,g., the immune system of an organism, are notably complicated in two ways: (1) they are tuned to respond differently to different system inputs and (2) the system that mediates outputs as a function of inputs is full of network redundancy, which ensures operation under non-ideal circumstances. Thus, learning the rules for how a complex system operates requires *coincident varied,* and likely, *combinatorial* external and internal perturbations to a system. Mild perturbations are likely buffered by the system; strong perturbations are the key. Complex systems often have numerous parts; how do we decide which parts should be perturbed? Network hubs—parts which integrate numerous signals and/or regulate many parts—are

excellent candidates for an abbreviated parts list. We then monitor system behavior when different combinations of hubs have been rendered inoperable. Such data can enable mathematical reconstruction of how major system parts are stimulated, influence each other, and modulate system output. Goals of these models include directing further experiments, furthering our fundamental understanding of the system, and predicting properties of the system relevant in applied settings. Image: conceptual diagram of a complex system responding to an external input (perturbation).

Chapter 3.
A search for mechanistic mathematical models of the JA-ET-PAD4-SA network.

3.1 Chapter Overview

A major goal of computational systems biology is to discover and quantitatively characterize the key mechanisms that underlie biological systems of interest. The purpose of such models is to be able to predict system behavior, even far from the conditions under which the data were collected for model fitting. We call such predictions type II predictions (Hillmer and Katagiri 2016); such predictions are the gold standard for models in the physical sciences. In this chapter, I present my work in attempting to discover predictive mechanistic mathematical models for the plant immune signaling network. This network can be approximated by four signaling sectors: the jasmonate (JA), ethylene (ET), phytoalexin-deficient 4 (PAD4), and salicylate (SA) signaling sectors (Tsuda et al. 2009; Kim et al. 2014).

I sought models that could explain observed dynamic modulations in the activities of all four sectors after stimulating the plant immune system with bacterial flagellin (flg22). The models I considered all assume that sector activities can be explained by two types of regulation: (1) network input(s), and (2) the activities of any of the four sectors. These are reasonable assumptions since the JA-ET-PAD4-SA network controls about 80% of the pattern-triggered-immunity (PTI) conferred by flg22 against the virulent plant pathogen *Pseudomonas syringae* pv. *tomato* DC3000 (Tsuda et al. 2009).

I discuss two modeling frameworks: (1) models fit to $\log_2$ data that I have called "nonlinear delay models with an enforced 2D input-output relationship", or ND2D models for short, and (2) delay differential equations (DDEs) fit to concentration values of sector activities. Both models were fit to time-interpolated data. While the ND2D models showed isolated cases of excellent performance, they could not explain dynamic sector behavior for all network states that were profiled. DDE models showed more promise. A

simple DDE model was found for the behavior of each isolated sector. Fairly elaborate DDE models appear to be required to explain full network behavior; a case study for the PAD4 sector is shown. This PAD4 model can explain the data well, although this model requires approximately as many parameters as there are mean data points underlying the data time-interpolation. Thus, while effective, this PAD4 model is potentially fitting time-interpolation artifacts as well as network behavior; additional work will be needed to determine to what extent this PAD4 model captures the network's underlying mechanisms. This chapter concludes with a discussion of what additional data and analyses may be useful in future work to discover mechanistic mathematical models of the Arabidopsis plant immune signaling network.

The modeling work done in this chapter was performed in close collaboration with my adviser, Fumi Katagiri, and with regular input from Chad Myers in the Department of Computer Science and Engineering at the University of Minnesota. The data used in this chapter were generated from flg22 infiltration experiments performed by Kenichi Tsuda, who was at the University of Minnesota at the time. Tag-Seq libraries (Rallapalli et al. 2014) for expression profiling generated from these plant tissue samples were mainly made by Kenichi Tsuda, Ghanashyam Rallapalli, and Shuta Asai, with some assistance from William Truman and myself; Ghanashyam's and Shuta's work was done in the laboratory of Jonathan Jones at The Sainsbury Laboratory in Norwich, UK; William's work was done at the University of Minnesota. Hormone profiles were generated by Hitoshi Sakakibara at the RIKEN Center for Sustainable Resource Science in Japan, with assistance from Mikiko Kojima.

3.2 Background: Coarse-grained empirical modeling of the Arabidopsis plant immune signaling network

To date, most mechanistic mathematical models with type II predictive power for dynamic systems in molecular systems biology are derived from Michaelis-Menten chemical kinetics. For large complex biological systems, however, such a modeling framework is impractical. The needed kinetic parameters when hundreds of parts are in play are mostly unknown, and the complexity of differential equations with more than a

handful independent variables rapidly becomes unwieldy. A potential alternative, successfully employed in empirical biological sciences such as ecology, epidemiology and engineering, is to directly discover, in a data-driven manner, the rules governing system dynamics at a more summarized scale (Ellner and Guckenheimer 2011). This summarization involves lumping numerous related parts into a single variable. Summarization aims to capture the essence of a system's behavior while glossing over minor variations in this behavior.

The models described in this chapter were preceded by two successful, descriptive models of the Arabidopsis immune signaling network (Tsuda et al. 2009; Kim et al. 2014). Tsuda et al. developed the genetic system that demonstrated that the JA-ET-PAD4-SA network approximates the Pattern-Triggered-Immunity (PTI) signaling network in Arabidopsis. Null mutants in the *DDE2*, *EIN2*, *PAD4*, and *SID2* genes were used to remove the JA, ET, PAD4 and SA sectors from the network, both individually and combinatorially via 4 single-gene mutants, 6 double-gene mutants, 4 triple-gene mutants, a quadruple mutant and the wild type. From PTI bacterial growth data collected in all these different genotypes, the contributions of individual sectors to immunity, as well as the immune contributions of two-sector, three-sector and four-sector interactions, were quantitatively estimated using a mixed-effects linear modeling framework which was called "signaling allocation analysis." This model captured causal genotype to phenotype relationships.

Statistical models that link input and output values serve as an important first step in modeling a biological system, but they lack mechanistic descriptions of network function. A limitation of the Tsuda et al. model was that it only explained network output (PTI), leaving the signal flows that produced this network output as a black box. The signaling allocations for immune phenotypes do not explain the network dynamics that lead to those network outputs. Rather, they are static descriptions of the network output that approximate the net effect of the dynamic signaling within the network. It should be noted that when signaling allocation analysis is applied to dynamic network response data, rather than static data, the mechanisms behind network dynamics can be investigated. In the next chapter, Chapter 4, I characterize the dynamics of the flg22-

responsive transcriptome by applying signaling allocation analysis to each gene's dynamic expression.

The natural next step after an initial set of statistical models is to open up the black box that links network input to output. We did this by modeling the dynamic signal flows in the JA-ET-PAD4-SA network that drive the PTI network output (Kim et al. 2014). From sector activity measurements via gene expression proxies at two time points after network stimulation by microbe-associated molecular patterns (MAMPs) as well as network output data (PTI bacterial growth data), we built a linear-regression-based semi-dynamic model of PTI, to which I contributed ideas for statistical analysis and model interpretation. This model is called semi-dynamic since two time points are the minimum needed to capture dynamics. This model revealed causal signal flows between sectors, and from sectors to network output. Notably, for this model we did not have to assume different signal flows in the network for different MAMPs; a single model structure explained MAMP-responsive signaling in the JA-ET-PAD4-SA network. However, the rules for these signal flows between sectors changed both quantitatively and qualitatively in time.

The next step towards predictive dynamic modeling of the plant immune signaling network is to find mechanistic rules that explain the dynamic signal flows in the network that were captured in the Kim model (Kim et al. 2014). By mechanistic rules, I mean fixed mathematical relationships for how the network sectors interact; mechanistic rules are themselves time-invariant, but they predict dynamic behavior. Such models are the modeling standard in the physical sciences. For example, the law of gravity, true everywhere in space as time that we know of, can be used to predict the orbital dynamics of planets. In the work described in this chapter, I sought mechanistic rules for sector interactions that could explain sector activity dynamics after flg22 treatment.

3.3 ND2D models of the JA-ET-PAD4-SA network

The models fit in this chapter were fit to detailed dynamic sector activity time course profiles collected 0, 1, 2, 3, 5, 9, and 18 hrs after flg22 treatment across the full set of

combinatorial genotypes used in both (Kim et al. 2014) and (Tsuda et al. 2009). These data were extracted from Tag-Seq (stranded mRNA) libraries and LC-MS hormone profiles (see Chapter 4 for details). Gene expression proxies (AT4G04500 for PAD4 and AT5G39900 for ET) were used to monitor ET and PAD4 sector activities; free JA and free SA hormone measurements were used for JA and SA sector activities.

3.3.1 Selection of sector activity marker genes

The sector activity marker genes used in the Kim model (Kim et al. 2014) were insufficient for two reasons: (1) the transcript levels of some of the marker genes measured by qPCR in that study (Kim et al. 2014) were below the detection limit of our Tag-Seq libraries and (2) the Tag-Seq dataset allowed me to implement far more conservative selection criteria for sector activity proxies than were possible to implement with the literature microarray datasets that were mined for the Kim et al study (Kim et al. 2014). Chapter 4 details my selection of the PAD4 and ET sector activity marker genes. Additional candidate sector activity marker genes for the PAD4 and ET sectors, which may be useful in future work, are also listed in Chapter 4.

3.3.2 Interpolation of sector activity data

A major feature of the Kim model (Kim et al. 2014) was the need for delays in sector cross-talk and for delays in sector activation. Thus, I expected delays to also play a prominent role in the models I fit for the work described in this chapter. Delays could in principle take on any continuous value, so I needed to interpolate our sector activity time course data so as to estimate sector activity at time points between measured data points. We determined that interpolation should have the following desirable features:

1. The interpolation should be a smooth curve, everywhere continuous and differentiable between data points (i.e., no gaps and no sharp edges).
2. The endpoints should be well-behaved. Specifically, the derivative should go to 0 at 0 hrs, since we assume that the system is responding from steady state.

For the models described in this section (non-linear delay models with an enforced 2-D linear input-output relationship, which I have named ND2D models), these features are

not strictly needed, but they are critical for the differential-equation-based models in section 3.4. Nonetheless, I chose to use the same interpolation for the models in both section 3.3 and section 3.4 to avoid arbitrary differences in input data between the two types of models.

We used the following data interpolation procedure to implement the desirable features above ((1), (2)). To ensure a smooth interpolation curve, polynomial regression was selected. Polynomial fits for short time course data perform better on the [-1, 1] interval than on the wider Real axis. Therefore we used a spline transformation to transform the sampled time interval, [0, 18] hrs, onto a pseudo time axis where we performed the polynomial fitting. The interpolated curve was returned to the real time axis via the inverse of this spline transformation.

To avoid the tendency of all non-constant polynomial terms to go to $y = 0$ at 0, we chose the sub-interval [-1,-0.5] (referred to as "pseudo time" hereafter) on the well-behaved interval [-1,1]. Since we required not only a smooth interpolation, but more than that, a smooth interpolation with well-behaved endpoints, we selected the following basis set for fitting:

$$y \sim (t + 0.5)^2 (t + 1)^2 f(t) \qquad \textbf{Equation 3-1.}$$

where t is the pseudo time variable, $(t + 0.5)^2$ is an envelope function that forces y to be twice-differentiable at $t = -0.5$, and $(t + 1)^2$ is an envelope function that forces y to be twice-differentiable at $t = -1$, and $f(t)$ is a polynomial, e.g. $a + bt + ct^2 + dt^3 + et^4$ where a, b, c, d, and e are constants. Mean steady state (0 hr) values of sector activity were subtracted off from each genotype separately so that a constant value was not needed in fitting; these constant values were added back on after fitting. Activity values for each sector were interpolated separately; only data from genotypes with the given sector present were interpolated. We assumed that the system starts from steady state at a real time of 0 hrs (pseudo time = -1) and that the system relaxes back to the same state at 36 hrs (pseudo time = -0.5). We assume that the system returns back to steady state a long time after stimulus, since it has been observed that the protective effects of

the flg22 stimulus peak after about a day and fully diminish over time. 36 hrs was an arbitrarily-chosen, but reasonably large, value that satisfied this known behavior. Thus we used 8 time points: 0, 1, 2, 3, 5, 9, 18, and 36 hrs, where the 0 hr data was copied at 36 hrs.

An unbiased fit would space the measured time points and the endpoint anchor (0, 1, 2, 3, 5, 9, 18, 36 hrs) evenly on the sub-interval [-1, -0.5] of the pseudo time axis. But we found that this produced a less-than-optimal fit. In intervals where there was little change in sector activity, the polynomial was too flexible and wiggled too much, whereas in intervals where there were large changes in data values (such as JA sector activity from 0 hr to 1 hr), the polynomial was insufficiently flexible enough and couldn't bend enough. Thus we spaced the real time points on the [-1, -0.5] pseudo time interval based on the absolute value of the discrete 2nd derivative. The continuous second derivative describes the curvature of a curve at any time point; the discrete second derivative approximates this curvature.

The discrete 2nd derivative was calculated as follows: for each genotype with the sector of interest present, the change between mean activity values at consecutive time points was calculated; this is the discrete 1st derivative. This yields a vector of values one index shorter than the total number of time points. An extra value of '0' was included at both the beginning and the end of this vector, since we assume steady state before and long after flg22 stimulus. The difference between consecutive values in the discrete derivative vector was calculated; this is the discrete 2nd derivative. The absolute value of the discrete second derivative was calculated; the changes between consecutive absolute discrete 2nd derivative values, divided by the time elapsed during that interval was used to calculate the relative size to be used for each time interval. The mean of these relative sizes across all relevant genotypes for each interval was calculated, and these 7 mean sizes, one for each interval, were scaled together so that the total distance was 0.5, the length of the pseudo time fitting interval.

The spacing I selected varied for the four different sectors, since the sector dynamics are markedly different for the different sectors. For example, JA rapidly responds from 0 hrs to 1 hr while SA is relatively constant on the same interval. To reduce uncertainty

in the time-interpolated mean estimates, a single polynomial fit was used for each sector (using genotype as an interacting factor with the polynomial), incorporating the same spacing across all nine genotypes with the given sector (wild type, 3 single mutants, 3 double mutants, 1 triple mutant and *fls2*). Figure 3-1 shows the time point spacings chosen and the spline transformations used for the data from each sector.

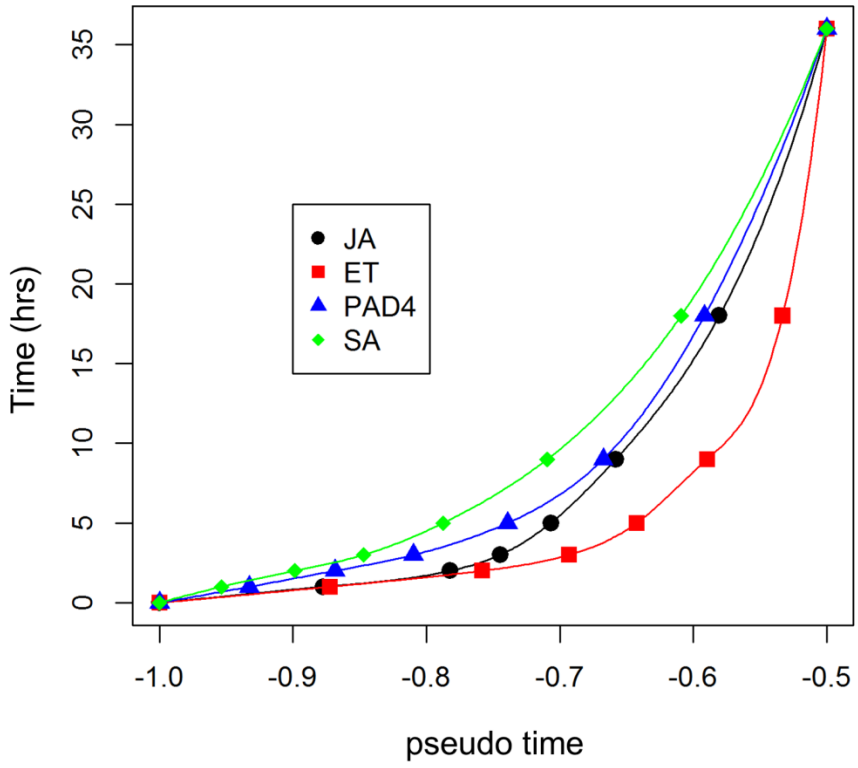


Figure 3-1. The time axis transformations used for polynomial interpolation.

When selecting the polynomial interpolations used in this chapter, fits with varying maximal polynomial degrees were visually inspected to guarantee interpolation curves that followed the underlying data points. For JA and PAD4, $f(t) \sim a + bt + ct^2 + dt^3 + et^4$ was selected. For ET and SA, $f(t) \sim a + bt + ct^2 + dt^3$ was selected. Pseudo time fits are shown in Figures 3-2 to 3-6.

Combinatorial genotypes are referred to in this chapter using single-letter abbreviations: J for JA, E for ET, P for PAD4 and S for SA. Lower-case letters are used to indicate that a sector has been removed in a given genotype. For example, *dde2 ein2* is referred to as *jePS*.

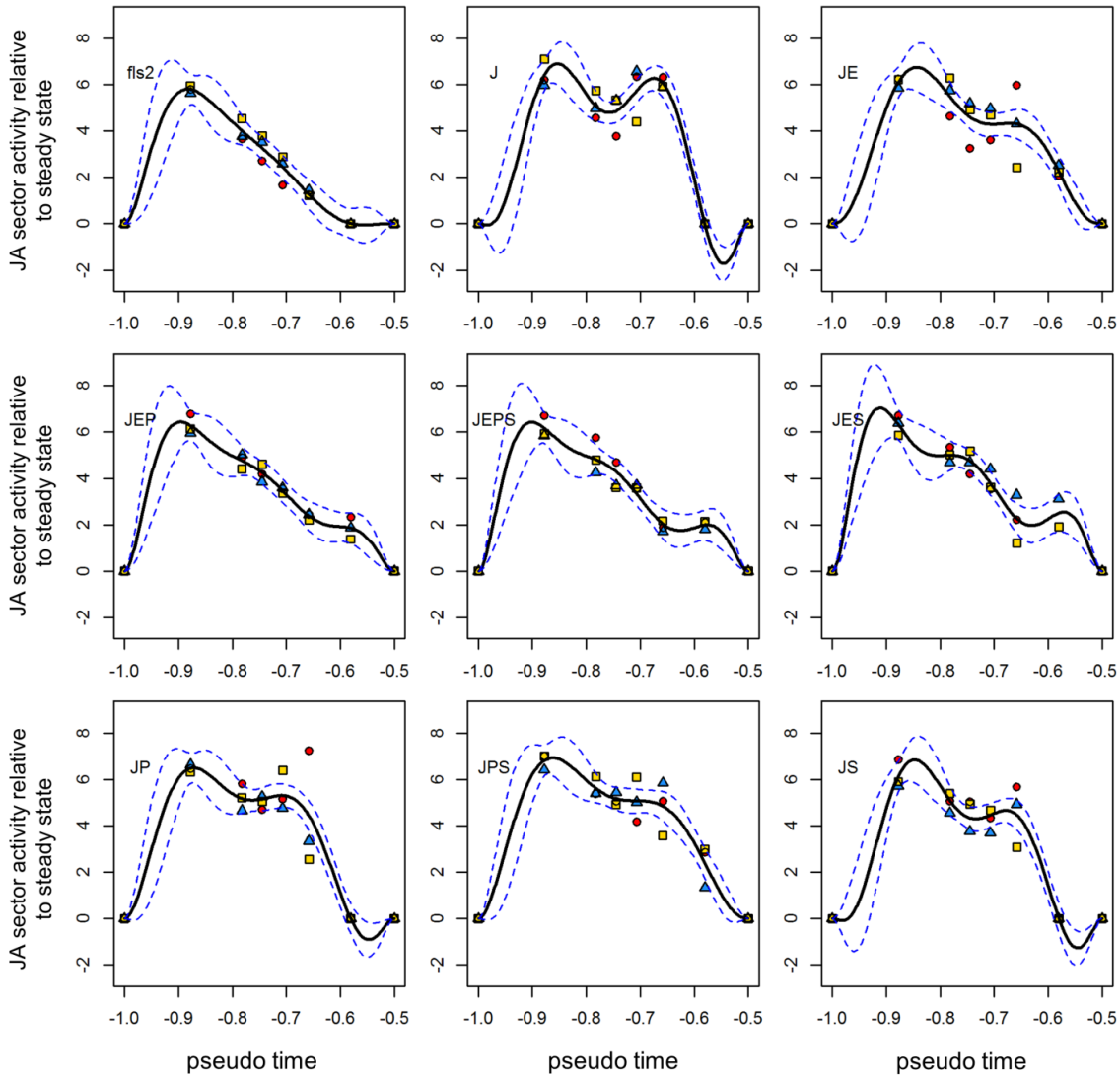


Figure 3-2. Pseudo time interpolations of JA sector activity data. Smooth black line: interpolated values. Dashed lines: 95% confidence intervals (CIs) for mean interpolated values. Symbols and colors indicate the biological replicate of the underlying data points: red circle, rep1; blue triangle, rep2; yellow square, rep3. For easy viewing, genotype abbreviations only indicate sectors present. For example, J is *Jeps*.

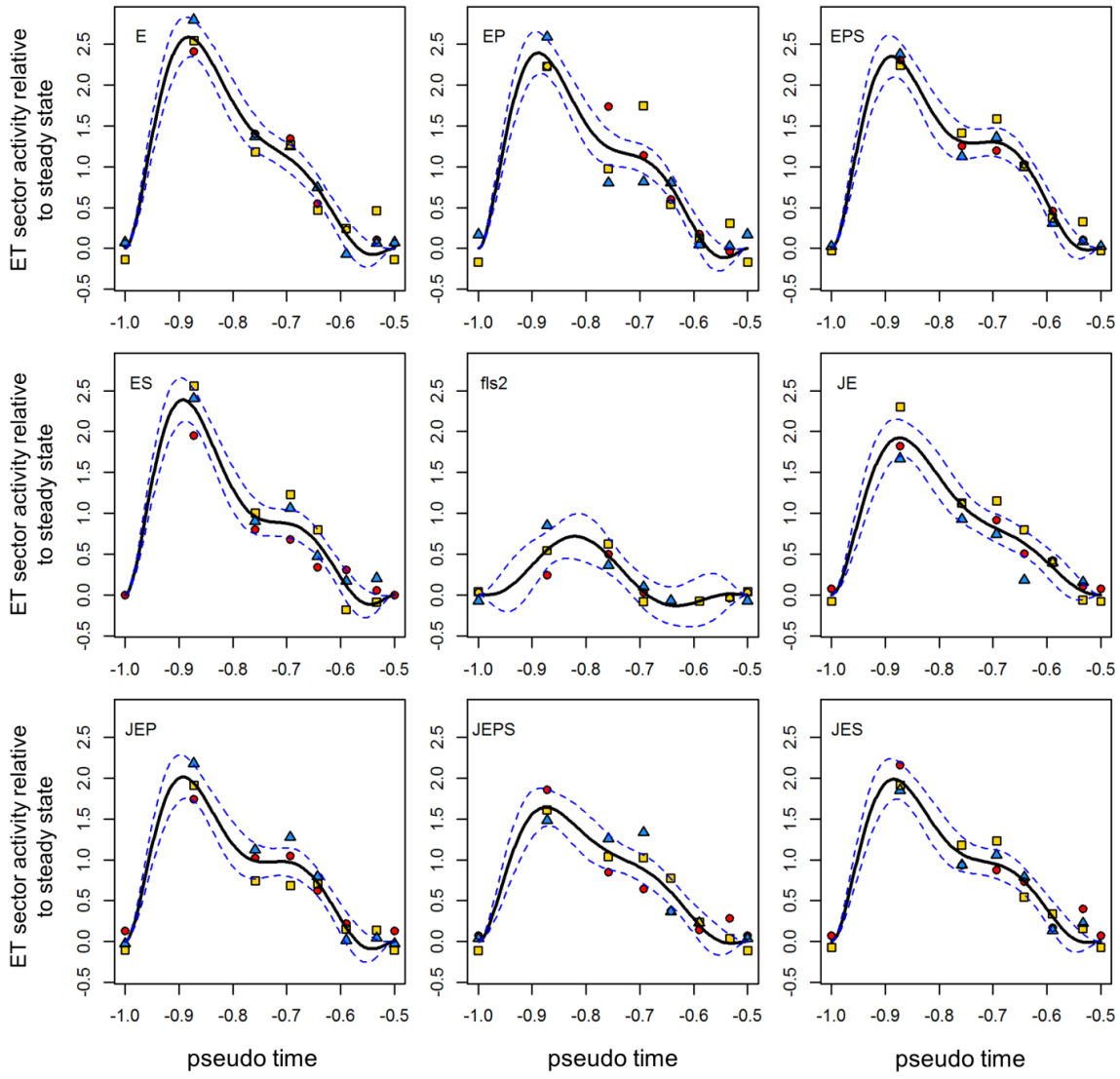


Figure 3-3. Pseudo time interpolations of ET sector activity data. Smooth black line: interpolated values. Dashed lines: 95% confidence intervals (CIs) for mean interpolated values. Symbols and colors indicate the biological replicate of the underlying data points: red circle, rep1; blue triangle, rep2; yellow square, rep3. For easy viewing, genotype abbreviations only indicate sectors present. For example, E is *jEps*.

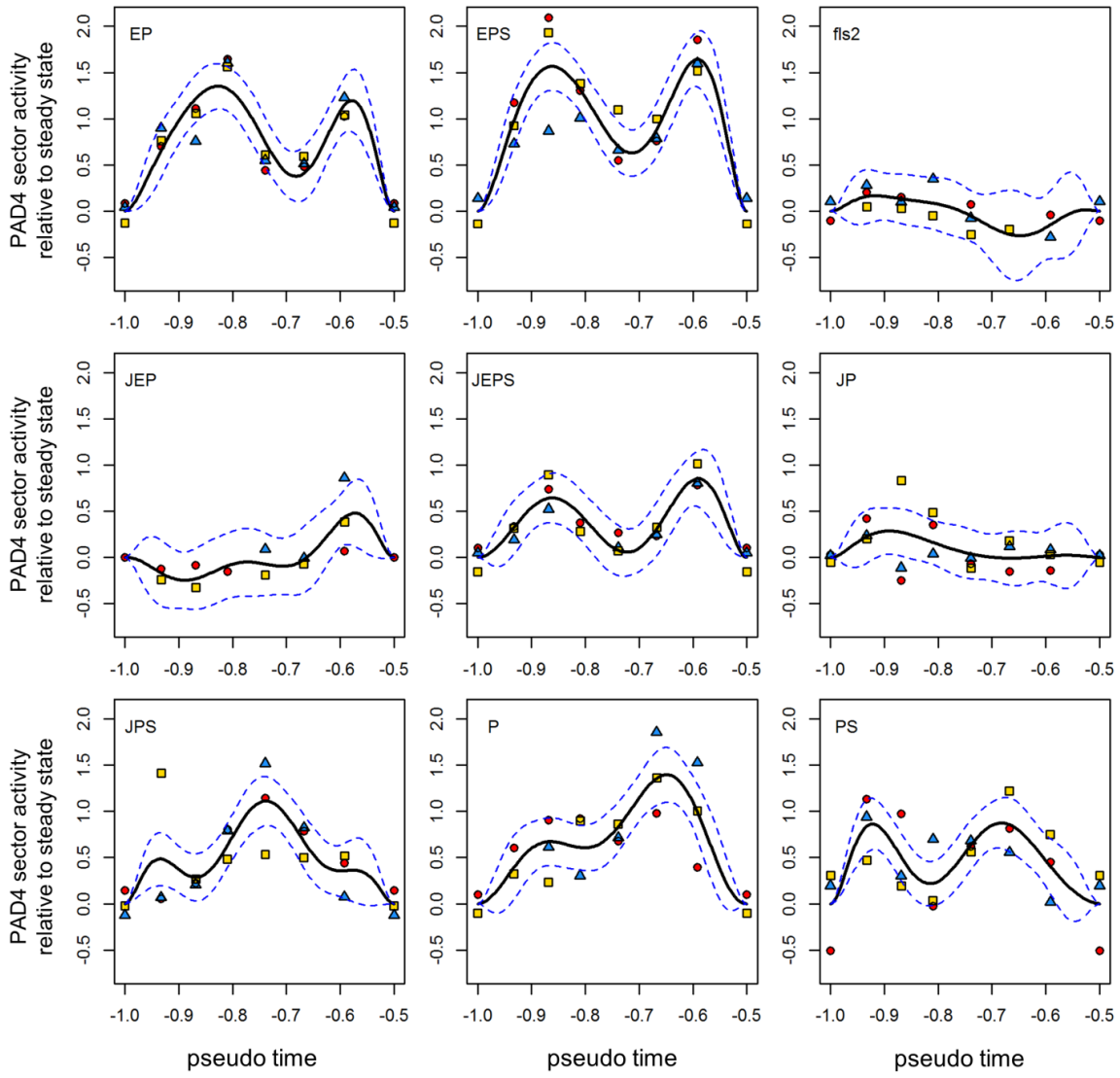


Figure 3-4. Pseudo time interpolations of PAD4 sector activity data. Smooth black line: interpolated values. Dashed lines: 95% confidence intervals (CIs) for mean interpolated values. Symbols and colors indicate the biological replicate of the underlying data points: red circle, rep1; blue triangle, rep2; yellow square, rep3. For easy viewing, genotype abbreviations only indicate sectors present. For example, P is *jePs*.

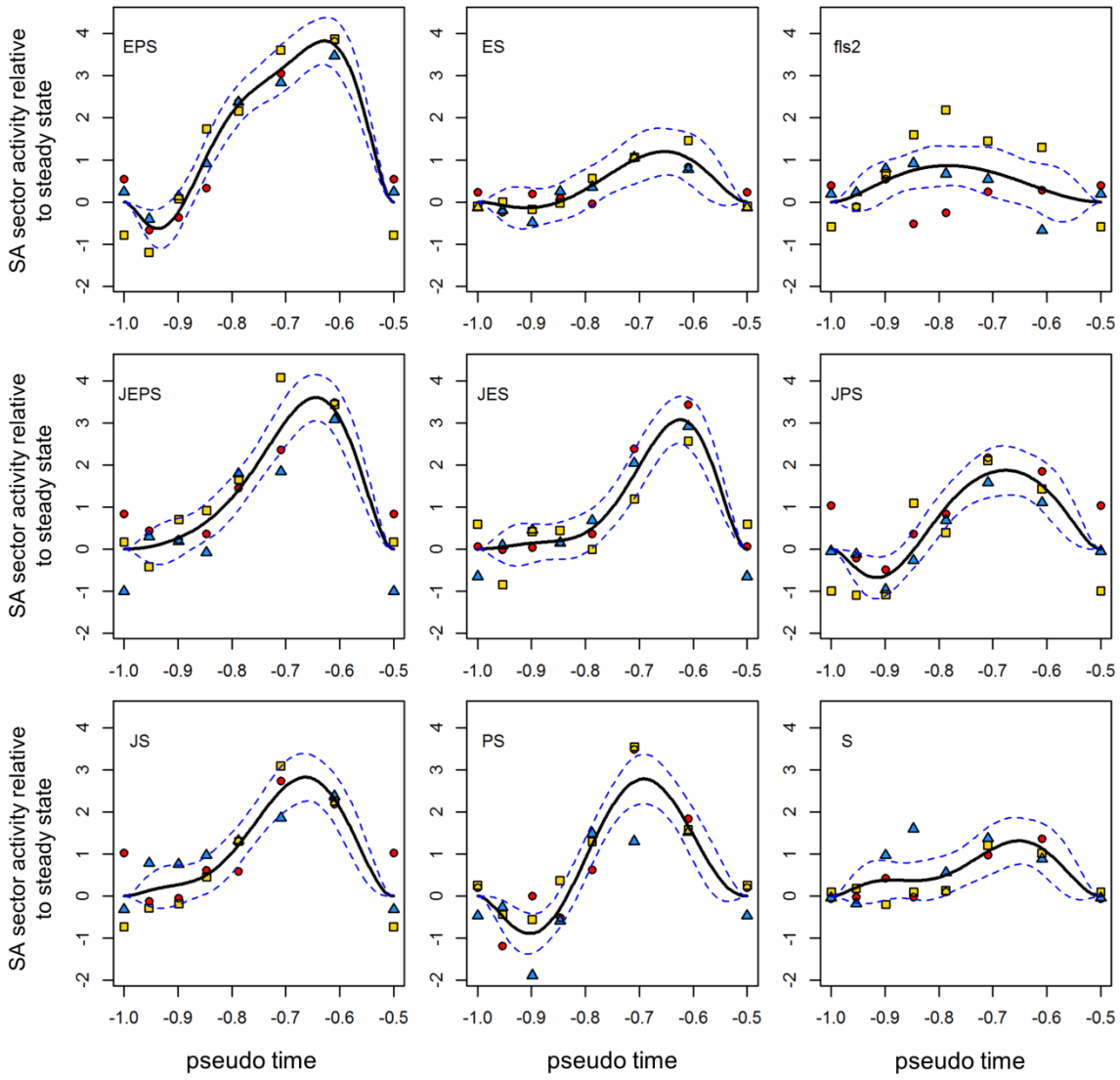


Figure 3-5. Pseudo time interpolations of SA sector activity data. Smooth black line: interpolated values. Dashed lines: 95% confidence intervals (CIs) for mean interpolated values. Symbols and colors indicate the biological replicate of the underlying data points: red circle, rep1; blue triangle, rep2; yellow square, rep3. For easy viewing, genotype abbreviations only indicate sectors present. For example, S is *jepS*.

To produce the sector activity curves used for the models in this chapter, pseudo time interpolations were transformed back to the [0, 36] hrs time interval using the inverse of the sector's spline time transformation (see Figure 3-1), and mean steady state values for each sector and genotype were added back on. For the ND2D models in section 3.3, interpolated curves were sampled every 0.2 hrs, an arbitrary but small value intended to provide a reasonable number of samples to capture the shape of the interpolated curve,

while limiting computational complexity when I searched time delays. Interpolated sector activity data for each sector are shown in Figures 3-6 to 3-9.

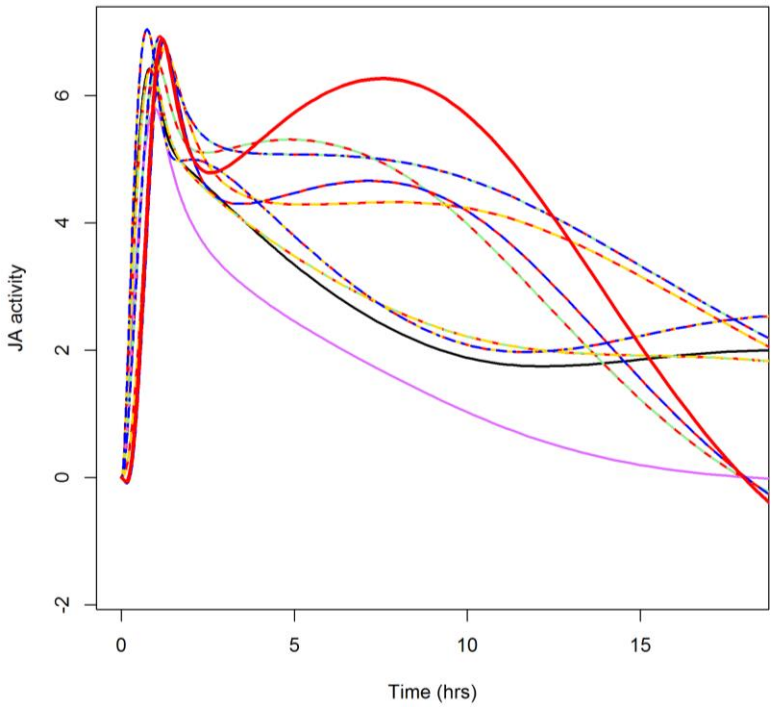


Figure 3-6. Time-interpolated JA sector activity values. Black curve, wild type; purple curve, *fls2*. Other genotypes are denoted using colors that indicate sector presence. Red, JA; yellow, ET; green, PAD4; blue, SA. For example, the blue and red curve is the interpolated JA time course for *JepS*.

Note that the slight differences in peak time for free JA concentration may be an artifact of our fitting procedure. The underlying mean activity data reach a maximum at 1 hr (among the sampled time points 0, 1, 2, 3, 5, 9, and 18 hrs), but the speed at which the sector activity in different genotypes fall from that peak affects the polynomial curvature near the 1 hr peak. However, if the rigidness of the polynomials used accurately reflects the underlying sector behavior, then these peak differences could be real. Collecting additional early JA sector activity measurements in the future will be necessary to accurately determine the JA sector behavior in the first hour after flg22 treatment.

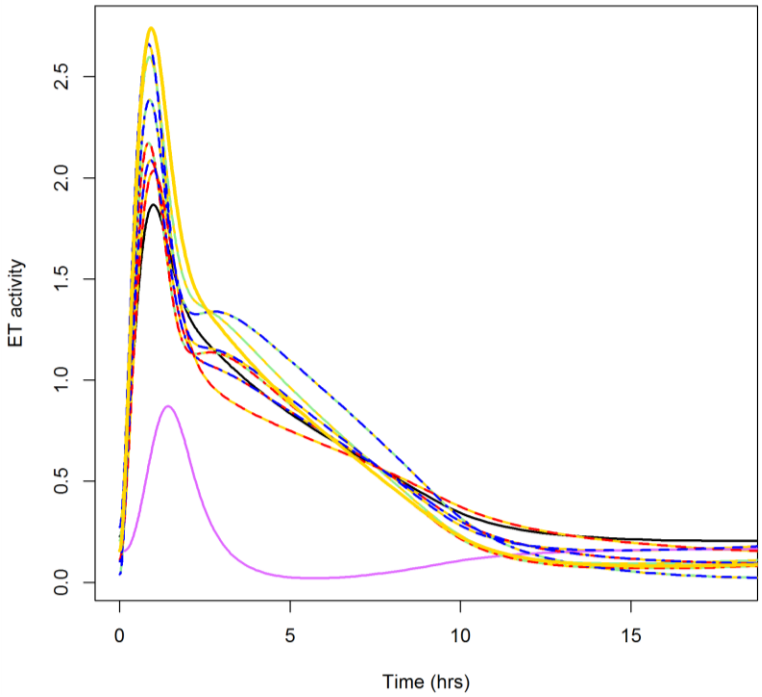


Figure 3-7. Time-interpolated ET sector activity values. Black curve, wild type; purple curve, *fls2*. Other genotypes are denoted using colors that indicate sector presence. Red, JA; yellow, ET; green, PAD4; blue, SA. For example, the yellow and green curve is the interpolated ET time course for *jEPs*.

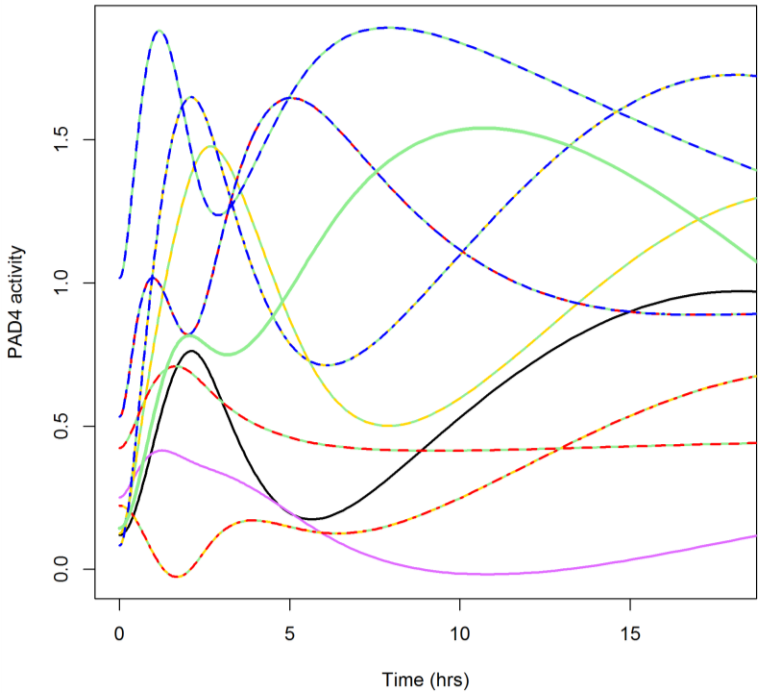


Figure 3-8. Time-interpolated PAD4 sector activity time courses. Black curve, wild type; purple curve, *fls2*. Other genotypes are denoted using colors that indicate sector presence. Red, JA; yellow, ET; green, PAD4; blue, SA. For example, the red and green curve is the interpolated PAD4 time course for *JePs*.

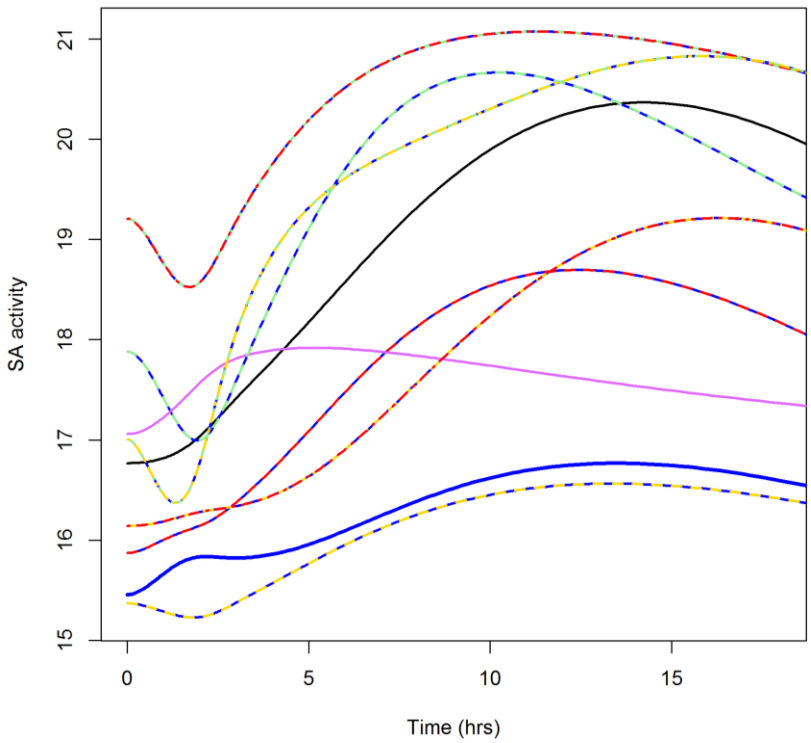


Figure 3-9. Time-interpolated SA sector activity time courses. Black curve, wild type; purple curve, *fls2*. Other genotypes are denoted using colors that indicate sector presence. Red, JA; yellow, ET; green, PAD4; blue, SA. For example, the yellow and blue curve is the interpolated SA time course for *jEpS*.

3.3.3 ND2D model structure discovery

With a good data interpolation in hand, I searched for mathematical model structures that could provide hypotheses about the mechanisms that could generate curves like those of the interpolated data. As noted in the introductory chapter, (Chapter 1), there are no fundamental rules from which to derive the underlying rules of the JA-ET-PAD4-SA network; the modeling scale at which Michaelis-Menten kinetics are applicable, where tested and reliable dynamic mechanistic models exist, is far too removed from the scale of the signaling sectors—JA, ET, PAD4, and SA. Therefore, we required a data-driven empirical modeling approach to discover the rules underlying the network de novo.

Note that an exhaustive search for mathematical functions with the ability to explain response variables as a function of input variables is a computationally impossible

search. It would involve searching an uncountably infinite set: the set of all mathematical functions known (and unknown) to man. Instead, we relied on the pattern-detection capabilities of the human brain to guess functional forms which might explain patterns in the data. Once a mathematical form was chosen, we searched for parameters to fit that model to the data. One mathematical form that showed promise was a form I have named, “nonlinear delay models with an enforced 2-dimensional input-output relationship”, or, hereafter, ND2D models.

A key constraint on the Kim model (Kim et al. 2014) was that sectors could not influence themselves: the mathematical undecidability between whether changes in sector activity came from influences from a MAMP to a sector, or via the auto-regulation of a sector, produced a singularity in the Kim model. Therefore, the models disallowed auto-regulation. A similar limitation arose with our Tag-Seq dataset. The quadruple mutant response serves as the network response baseline. In a triple mutant two things have been added to the system: both the sector of interest with any auto-feedback it may have, as well as any input from flg22 to that sector. The shape of the flg22 input to each sector is unknown as of this writing. Some details about receptor-mediated endocytosis of flg22 bound to its cognate receptor, *FLS2* have been measured: GFP-labeled receptors appear to have been mostly internalized within 20-40 minutes after flg22 treatment (Robatzek, Chinchilla, and Boller 2006). Yet the details of how this signal is transduced to each sector are unknown. They include the possibility that internalized receptors may continue engaging in intra-cellular signaling. In addition, there could be signaling steps between the receptor and the sector in question, which delay the signal and/or modify the time-course shape of the signal before the signal reaches to the sector. Thus, given uncertainty regarding the network input signals, the ND2D models take the triple-mutant behavior as given, without attempting to explain it. Rather, I sought to explain the response of mutant genotypes with more intact network states, starting with the double-gene mutants. In each double mutant, a single factor has been added to the system, relative to the triple mutant: one other sector. The ND2D models seek to explain the subtractive difference between wt, single or double mutant sector activities and the triple mutant activity of that sector. For example, for SA sector activity in the

double mutant *jePS*, I sought for models that could predict the time course values $SA_{\mathrm{jePS}}(\mathrm{t}) - SA_{\mathrm{jeps}}(t)$.

Via exploratory analysis, I noticed that in some cases, the differences between double and triple mutant sector activities roughly resembled the shape of the activity of the added sector. This similarity increased if I allowed delays between the added sector and the sector of interest. This suggested a mathematical rule structure whereby sector activities produce changes in other sector activities, with some delay. There are very plausible biological reasons for these delays: signaling from one sector to another could (and likely does) include time-consuming processes such as gene transcription, protein translation, protein maturation and vesicle trafficking. Delays on the order of minutes to several hours are reasonable features of any modeling scheme involving the summarized sectors JA, ET, PAD4, and SA.

If, in addition to a delayed input from the added sector, I added a multiplicative term between the sector of interest and the activity of the added sector, I could produce an extremely well-fitting model for the PAD4 effect on SA in the double mutant *jePS* (Figure 3-10).

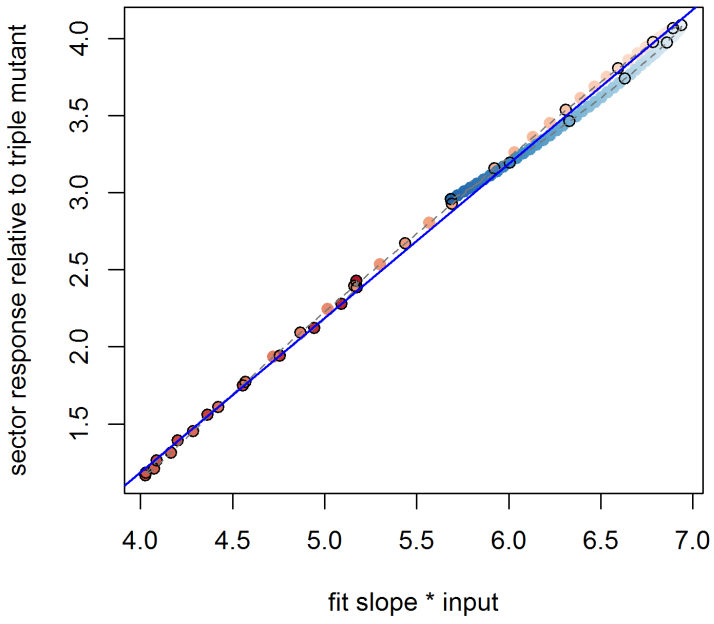


Figure 3-10. A very well-behaved input rule was found for the PAD4 effect on the SA sector in the double mutant *jePS*. (See Equations 3-2 and 3-3.) The adjusted R^2 value for the linear fit (Equation 3-3) to mean interpolated data was 0.997. The diverging color scheme

maps the implicit time dependence from the 0 hr time point (dark red) to the 18 hr time point (dark blue), through off-white at 9 hrs. Points are shown for all sampled times (every 0.2 hrs). Dark borders indicate the sub-sampled time points used in fitting. Note that the fit, with a slope of 1, will not go through the origin; the fit line contains an intercept. My ND2D models allow these flexible offsets. An interpretation of this intercept is that the PAD4 effect on the SA sector can be either positive or negative, depending on the sector activity ranges of the PAD4 and SA sectors.

The rule

$$input(t) = PAD4_{\mathrm{jePS}}(t + 1hr) * SA_{\mathrm{jePS}}(t) \; - 10 * PAD4_{\mathrm{jePS}}(t + 1hr)$$

Equation 3-2.

produced a very nearly linear relationship between the input (Equation 3-2) and the SA response, $SA_{\mathrm{jePS}}(\mathrm{t}) - SA_{\mathrm{jepS}}(t)$; $PAD4_{\mathrm{jePS}}(t)$ is the PAD4 activity in the genotype *jePS* as a function of time, and $SA_{\mathrm{jePS}}(t)$ is the SA activity in the genotype *jePS* as a function of time. That is, a linear model of the form

$$SA_{\mathrm{jePS}}(\mathrm{t}) - SA_{\mathrm{jepS}}(t) = a * input(t) + b$$

Equation 3-3.

fit the data well. During initial rule discovery, I optimized model parameters (e.g. +1hr, -10 in Equation 3-2) by hand.

Notice that there is a delay in the PAD4 sector activity values of -1 hr (Equation 3-2). The delays in this model have two different potential biological sources. The first is a delay in signaling between sectors. The second is a delay in the response of the sector activity proxy. The latter is included in the model as a negative delay, since the actual time for the sector activity may be ahead of the time when the sector activity proxy level was measured (both PAD4 and ET sector activities were measured via transcript levels of sector activity marker genes). For the ET proxy, any delay is likely quite short: using data in the Arabidopsis Electronic Fluorscent Pictograph (eFP) browser (D. Winter et al. 2007), the selected ET proxy responds strongly within 30 minutes after exogenous ACC treatment (ACC is an ethylene biosynthetic precursor), and very strongly 1 hr after treatment with the elicitors flg22 and HrpZ. For the PAD4 proxy, however, the only

information is (from our data) that a response occurs at 1 hour (the earliest time point in our dataset); in the eFP browser data, no strong responses are seen after any elicitor treatment at 1 hr or 4 hrs for our PAD4 sector activity proxy. It is possible that a delay of up to one hour could exist in our PAD4 sector activity marker, AT4G04500. Thus when PAD4 was the influencing sector, I searched both positive delays, as well as negative delays as negative as -1 hr. I only allowed negative delays as negative as -0.2 hrs for influences from the ET sector.

After time-interpolation, I sampled the interpolated data every 0.2 hrs. However, using all of these points in fitting would have been inappropriate. There are, for example, the same number of underlying real data points on the time intervals [0, 3] hrs and [3, 18] hrs. Therefore I sub-sampled the time-interpolated data relative to the underlying data density: six sample points, including the endpoints, were used on the [0, 1], [1, 2], [2, 3], [3, 5], [5, 9], and [9, 18] hrs intervals. This sub-sampled data was used to calculate adjusted R^2 values reported for the ND2D models in this chapter.

3.3.4 ND2D model structure and parameter selection for all double mutants

Via by-hand parameter exploration, I found that a set of related input rule structures appeared to be able to explain the effects of sectors on other sectors across all double mutants. I then fit all of these model structures to data using an exhaustive coarse-grained parameter search (Table 3-1). For cartoon diagrams of the full family of models considered, see Figure 3-11.

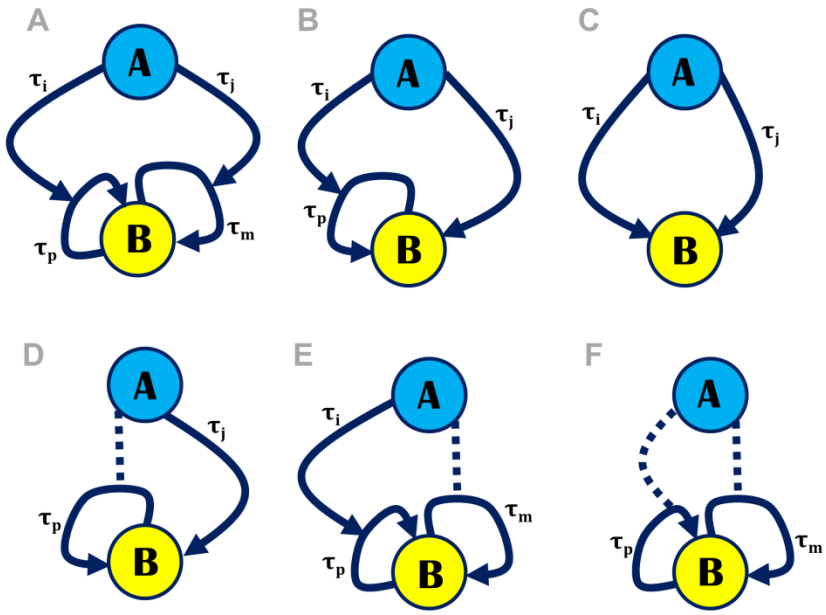


Figure 3-11. Diagrams of the family of ND2D models that I considered when searching for rules to explain double-mutant sector activity behavior. The A nodes represent the added sector in the double mutant, and the B nodes the sector of interest whose activity (relative to the triple mutant) is to be explained. The selected rule for the PAD4 effect on the SA sector (see Figure 3-10) has the structure B. τ_i, τ_j, τ_p, and τ_m are delays. Dotted lines indicate that the presence of sector A is required to activate the auto-feedback loop of B. Mathematically, the input rule for models of structure D, for example, take the form $input \sim A(t-\tau_j) + k * B(t-\tau_p)$. The relative coefficient k is not visually represented in the cartoon diagrams, but was searched during model selection. See Equations 3-4 (A-F) on page 34 for the mathematical expressions for all rule structures.

Fitting these ND2D models is not trivial. Unlike least-squares regression for ordinary linear models, there is no straightforward optimization procedure for ND2D models that I could find; fitting these ND2D models requires a computationally intensive global parameter search. Input rule selection involves exhaustively searching through delays and values for the relative coefficient which together produce a 1-1 input-output relationship.

By manually exploring parameter values, it appeared that some portions of some input-response curves could not be fit well, regardless of parameter values. Thus, for automated global parameter searches, I excluded the first several points of some response curves. Specifically, I removed the first 2 sampled points from the ET→SA and SA→ET rules, the first 3 sampled points from the SA→PAD4 rule, and the first 4

sampled points from the ET→PAD4 rule. Reported adjusted R^2 values do not include these points.

I performed an automated coarse-grained global parameter search (see Table 3-1) to find good models for all time-interpolated double-mutant sector activity data. This coarse-grained search allowed me to get a sense of how many different parameter sets could produce similarly well-performing models. It is very possible that these models could be carefully tuned to locally optimize fits (using a fine-grained parameter search instead of a coarse-grained one). But I expect that if anything, double mutant models are being over-fit, that is, fitting so closely to the interpolated data that interpolation artifacts are potentially being captured; therefore I was not interested in fine-tuning model parameter sets. The parameter values searched exhaustively are listed in Table 3-1 for the parameters i, j, p, m, and k, where i and j are the delays in the added sector, p and m are the delays in the sector of interest, and k is a coefficient. The full model (cartoon diagram A in Figure 3-10) uses all these parameters:

$$A(t - \tau_i) * B(t - \tau_p) + k * A(t - \tau_j) * B(t - \tau_m) \qquad \textbf{Equation 3-4A.}$$

Models with a subset of these available parameters were also searched (Figure 3-11 models B-F):

$$A(t - \tau_i) * B(t - \tau_p) + k * A(t - \tau_j) \qquad \textbf{Equation 3-4B.}$$

$$A(t - \tau_i) + k * A(t - \tau_j) \qquad \textbf{Equation 3-4C.}$$

$$B(t - \tau_p) + k * A(t - \tau_j) \qquad \textbf{Equation 3-4D.}$$

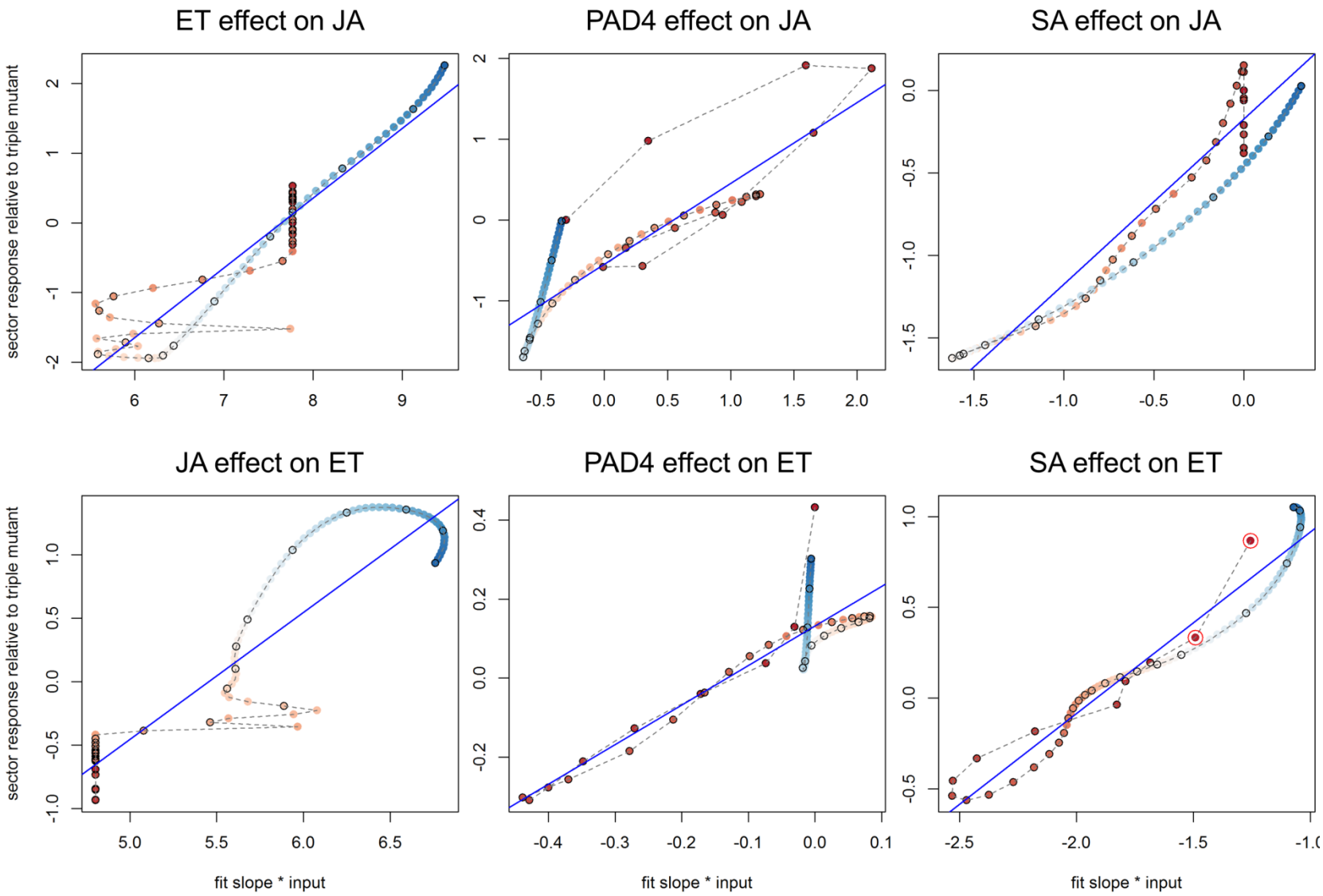
$$A(t - \tau_i) * B(t - \tau_p) + k * B(t - \tau_m) \qquad \textbf{Equation 3-4E.}$$

$$B(t - \tau_p) + k * B(t - \tau_m) \qquad \textbf{Equation 3-4F.}$$

Parameter	Values
i	JA, SA: [0,6] in steps of 0.4 ET: [-0.2,6] in steps of 0.4 PAD4: [-1,6] in steps of 0.4
j	JA, SA: [0,6] in steps of 0.4 ET: [-0.2,6] in steps of 0.4 PAD4: [-1,6] in steps of 0.4
p	[0,6] in steps of 0.4
m	[0,6] in steps of 0.4
k	[-10, -5, -3, -2, -1, -1/2, -1/3, -1/5, -1/10, 1/10, 1/5, 1/3, 1/2, 1, 2, 3, 5, 10]

Table 3-1. Coarse-grained model search parameter values. All possible parameter value combinations were evaluated. The upper range of delay parameters (6 hrs for all sectors) was chosen so as to be large, but not too large, since the underlying data density drops off after the 5 hr time point. The delay parameter density was chosen so as to be more dense than the density of the underlying data, but sparse enough to limit the computational load of exhaustively searching parameter space. Values of the coefficient k were chosen so as to allow both negative and positive combinations of terms in the input model, as well as up to an order of magnitude difference in weighting between the two terms in the input.

Good models were found for many sector effects on other sectors using the double-mutant time-interpolated sector activity data. See Figure 3-12 for input fit curves for the best performing model for each sector relationship, and Table 3-2 for the adjusted R^2 values of these fits.


ET effect on JA
PAD4 effect on JA
SA effect on JA
JA effect on ET
PAD4 effect on ET
SA effect on ET
sector response relative to triple mutant
sector response relative to triple mutant
fit slope * input
fit slope * input
fit slope * input

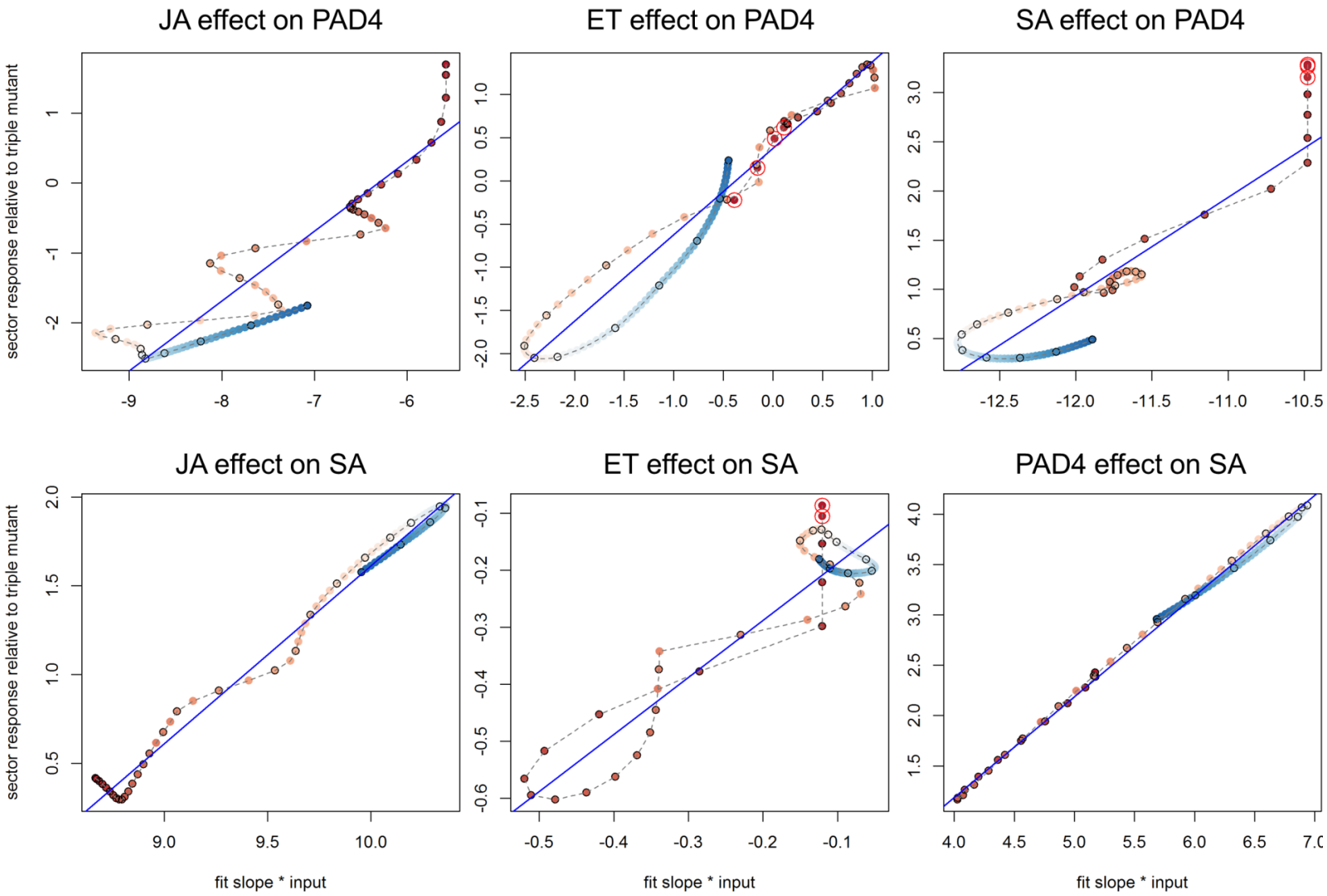

JA effect on PAD4
ET effect on PAD4
SA effect on PAD4
JA effect on SA
ET effect on SA
PAD4 effect on SA
sector response relative to triple mutant
sector response relative to triple mutant
fit slope * input
fit slope * input
fit slope * input

Figure 3-12. (see previous 2 pages) **Best-fitting ND2D model for each 2-sector relationship, fit to time-interpolated double mutant and triple mutant sector activity data.** The diverging color scheme maps the implicit time dependence from the 0 hr time point (dark red) to the 18 hr time point (dark blue), via off-white at 9 hrs. Points are shown for all sampled times (every 0.2 hrs). Dark borders indicate sub-sampled time points used in fitting. Dotted lines connect data points. Note that the fits, each with a slope of 1, do not go through the origin; the fit lines each contain an intercept. My ND2D models allow these flexible offsets. An interpretation of these intercepts is that model effects can be either positive or negative, depending on the sector activity ranges of the added sector and the sector of interest. Also note that the fit slope a can be negative; I multiplied by this slope (fit slope * input) so that a perfect fit shows up as a diagonal line with slope=1, to improve visual comparisons of rule performance.

Influence	i delay (hrs)	j delay (hrs)	p delay (hrs)	m delay (hrs)	k	Adjusted R^2
ET→JA	5.4	NA	3.4	5.0	-10	0.870
PAD4→JA	1.0	NA	0.2	0.6	-5	0.762
SA→JA	4.2	1.8	2.6	2.6	-1	0.910
JA→ET	NA	4.6	5.0	NA	-1	0.872
PAD4→ET	1.8	0.2	0.6	0.6	-1	0.818
SA→ET	2.6	NA	1.0	0.2	10	0.949
JA→PAD4	6.2	3.8	1.0	0.6	1	0.865
ET→PAD4	0.2	3.8	1.4	5.0	-2	0.966
SA→PAD4	NA	NA	1.4	3.0	0.5	0.881
JA→SA	NA	3.8	0.2	NA	0.1	0.984
ET→SA	1.0	3.8	5.0	NA	-10	0.888
PAD4→SA	-1.0	-1.0	0.2	NA	-10	0.998

Table 3-2. Parameters for the best fitting model for each sector's effect on every other sector's activity. Models were fit using mean interpolated double mutant and triple mutant sector activity data.

3.3.5 ND2D example rule composition: JA→SA and PAD4→SA.

Having found ND2D models that successfully fit the interpolated double mutant data, I sought to find composition rules which could explain the interpolated single mutant data. A case of interest is how the JA→SA and PAD4→SA rules combine. This was one

feature that the Kim model (Kim et al. 2014) could not explain. The data used to build the Kim model (Kim et al. 2014) showed non-linearity in how the JA and PAD4 effects on SA interacted when they were combined. But this non-linearity was not captured in the model, since a linear modeling framework was used for the Kim model. In adopting a more complex modeling framework, a litmus test is being able to capture this JA & PAD4 → SA behavior.

Surprisingly, I found a very simple rule that can explain the *JePS* SA data: dominance of the PAD4→SA rule. An ND2D model with different parameters and a different structure than those which produced the best-fitting double mutant model produced a single rule that could explain both *jePS* and *JePS* data (see Figure 3-13). (For how the best-fitting double mutant model fit the *JePS* data, see Figure 3-14.) An interpretation of this dominance as a composition rule is that the JA activation of SA is a backup effect that only switches on when the PAD4 activation is absent.

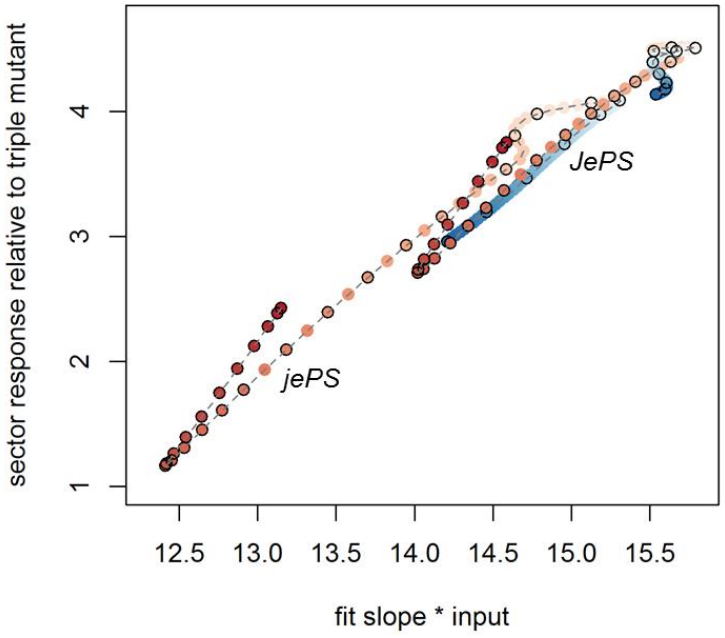


Figure 3-13. A single PAD4→SA model structure and parameter set can explain SA sector responses in both *jePS* and *JePS* relative to SA in *jepS*. Model parameters: i=NA, j=6.2 hrs, p=0.2 hrs, m=NA, k=-0.333. Rule selection: the minimum adjusted R^2 of the two model fits for the *jePS* and *JePS* responses was collected for each set of searched parameters (see Table 3-2 for parameters searched). The set of parameters with the maximal value for this minimum was selected. The diverging color scheme maps the implicit time dependence from the 0 hr time point (dark red) to the 18 hr time point (dark blue). Points are shown for all sampled

times (every 0.2 hrs). Dark borders indicate sub-sampled time points used in fitting. Dotted lines connect interpolated data points.

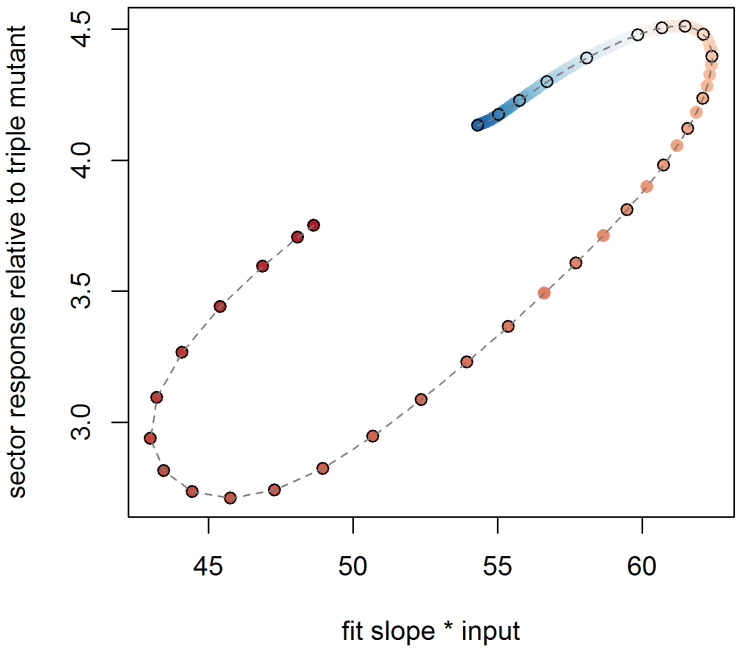


Figure 3-14. Performance of *JePS* data using the best fitting PAD4→SA *jePS* model.

Further, I investigated whether dominance of a PAD4→SA rule could explain SA sector behavior in all genotypes with active PAD4 and SA sectors: *jePS*, *jEPS*, *JePS*, and *JEPS*. Among all searched parameters, there was indeed a PAD4→SA model that could explain the SA sector response relative to the triple mutant *jepS* in all four genotypes. This rule fit to data from all four genotypes is seen in Figure 3-15. The parameters of this model were very similar to those for the best model for both *jePS* and *JePS* (Figure 3-13): only the k parameter changed from -1/3 to -1/5. One interesting feature of the fits in Figure 3-15 is that there appears to be an offset between the *EIN2*-containing genotypes and the *ein2*-containing genotypes. A possible interpretation of this offset is that the ET sector decreases the sensitivity of the SA sector to PAD4, since an offset in log-space equates to a multiplicative factor in concentration.

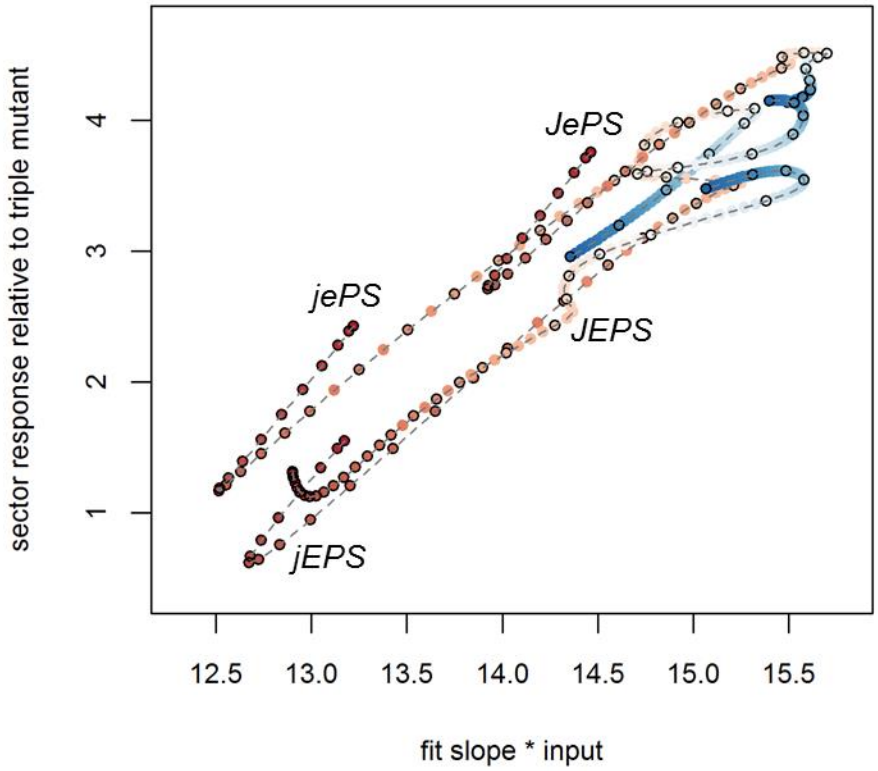


Figure 3-15. A single PAD4→SA rule can explain SA sector responses in all genotypes where PAD4 and SA are active. Model parameters: i=NA, j=6.2 hrs, p=0.2 hrs, m=NA, k=-0.2. Rule selection: the minimum adjusted R^2 of the four fits for the *jePS, JePS*, *jEPS*, and *JEPS* responses was collected for each set of searched parameters (see Table 3-2 for parameters searched). The set of parameters with the maximal value for this minimum was selected. Note the offset between genotypes with wild type and mutant ET sectors. The diverging color scheme maps the implicit time dependence from the 0 hr time point (dark red) to the 18 hr time point (dark blue). Points are shown for all sampled times (every 0.2 hrs). Dark borders indicate sub-sampled time points used in fitting. Dotted lines connect interpolated data points.

3.3.6 Additive combinations of ND2D models

While dominance proved a good rule composition framework for models that included the P→S interaction, dominance alone could not explain rule composition for all sectors. I therefore assessed an additive combination framework of ND2D models. In this framework, the double-mutant rules set the slopes (the a parameter in Equation 3-3), and the full rule is the sum of the individual rules.

It should be noted that often many different sets of parameters produced fits to double mutant data that were nearly as good as each other. I didn't want to only test additive

combinations of ND2D models using the best individual models. Rather, to test the performance of an additive combination framework for ND2D models, I searched all parameter sets whose double mutant adjusted R^2 values were within 1-3% of the best (maximal) adjusted R^2 value (1-3% was selected to limit computational load). I used 1% for the JA→SA, ET→SA, and PAD4→SA rules, where 3% collected too many models; for all other cases I used 3%. Table 3-3 shows the best performance of additive rule combinations for all single mutants and the wild type, for all four sectors. Note that the coefficients for these additive combinations are set in the double-mutant (the a parameter in Equation 3-3); this is not linear combination in the usual sense where the relative coefficients are fit using all the data. Rather, here the single mutant and wild type data serves as test data, which was held out from the initial model set selection.

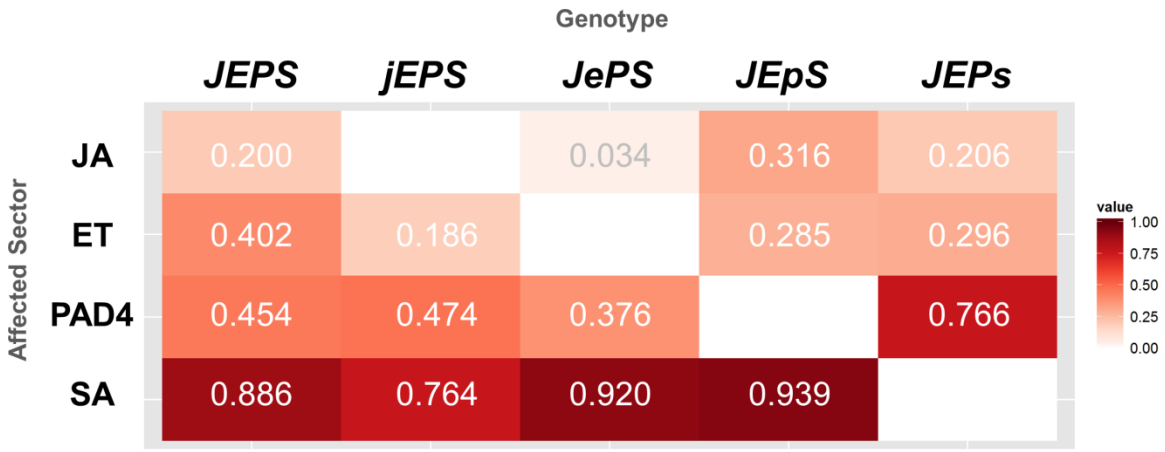


Table 3-3. Best adjusted R^2 values for additive combinations of ND2D models in wild type and all single-gene mutants.

Some additive combination of ND2D models worked reasonably well for fitting single mutant and wild type sector activity data (see Table 3-3; five combinations had Pearson correlations > 0.75 between interpolated data and modeled values). Other models showed poorer performance, notably attempts to explain either the ET sector activity data, or the JA sector activity data. The PAD4 sector activity in the *JEPs* genotype was explained well, but the other genotypes are only moderately well-fit. The SA sector activity data, on the other hand, was explained rather well. It is worth noting that the SA

sector activity data in the triple mutant *jepS* is the least dynamic among triple mutant activities for the four sectors (see Figure 3-9, compared with Figures 3-6 to 3-8). Explaining sector activity differences from the triple mutant using auto-regulatory terms (Equations 3-4D, 3-4E, and 3-4F) is meaningful when the triple mutant is dynamic, since then these differences are clearly distinct from the sector activity values being modeled. This distinguishability is less true for SA than for JA, ET and PAD4.

While the ND2D model framework showed some promise, it could not globally explain all sectors, under all perturbations. Moreover, the ND2D framework is non-standard; it was invented for the purpose of explaining our data. The hurdle of justifying a complicated novel modeling framework which produced less-than-excellent model performance spurred us in a new direction (see section 3-4). We next took a standard modeling framework for dynamic systems, differential equations, and looked to explain sector activity data.

3.4 A differential-equation-based framework for modeling the JA-ET-PAD4-SA network.

A second modeling framework that showed promise in capturing dynamic sector behavior was a set of mathematical models based on delay differential equations (DDEs). These DDE models describe the rate of change of a given sector as a function of inputs from outside the network, the sector's own activity at an earlier time, as well as the activities of other sectors, with variable delays.

3.4.1 Calculating concentration-based sector activity values and derivatives

When differential equations are used to describe chemical reactions, as in Michaelis-Menten chemical kinetics, the underlying variables are concentrations of different molecular species. Even though our sectors are highly summarized, their activities were measurable by concentrations of either free hormones (SA and JA) or transcript levels of sector activity proxies (ET and PAD4). We therefore translated the $\log_2$-based time-interpolations of sector activity from section 3.3 into concentrations (i.e., non-log

transformed expression or abundance values). We also calculated sector activity derivatives from these mean concentration and mRNA expression level estimates (Figure 3-15 and 3-16) to yield the values to which we fit differential equation models. SA data was re-interpolated using all 17 genotypes since basal SA is present in *sid2*-containing genotypes, and this basal SA was used for a DDE model of PAD4 sector activity. Most of this basal level of SA in *sid2* mutant backgrounds is known to be produced by Isochorismate Synthase 2 (ICS2), which is a close homolog of SID2 (ICS1) (Garcion et al. 2008).

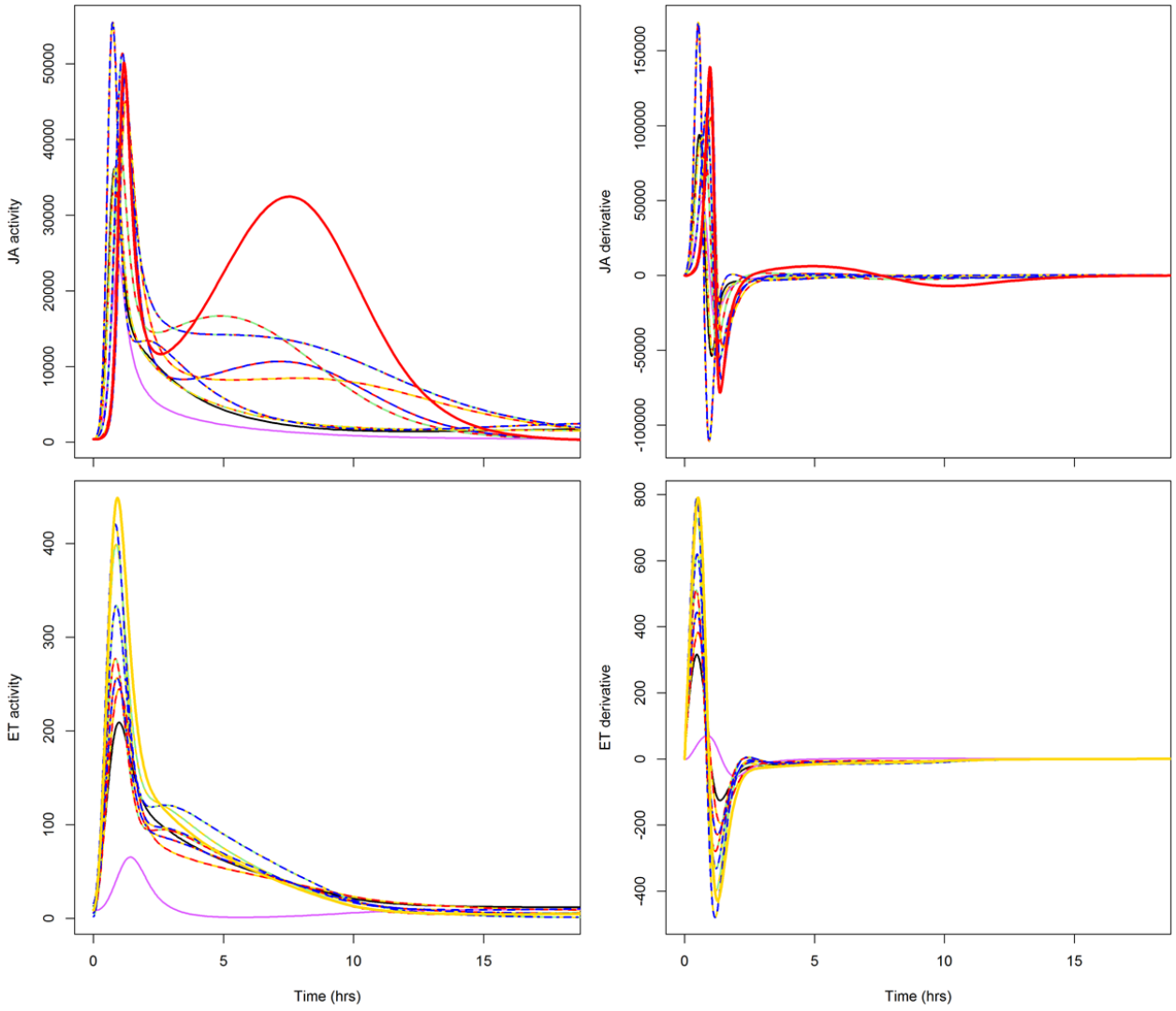


Figure 3-16. Time-interpolated JA and ET sector activity data and their derivatives, in concentration units (y-axis). Left: Activity; Right: Derivatives; Upper: JA; Lower: ET. Black curve, wild type; purple curve, *fls2*. Other genotypes denoted using colors that indicate sector

presence. Red, JA; yellow, ET; green, PAD4; blue, SA. For example, the red and yellow curve is the interpolated time course for *JEps*.

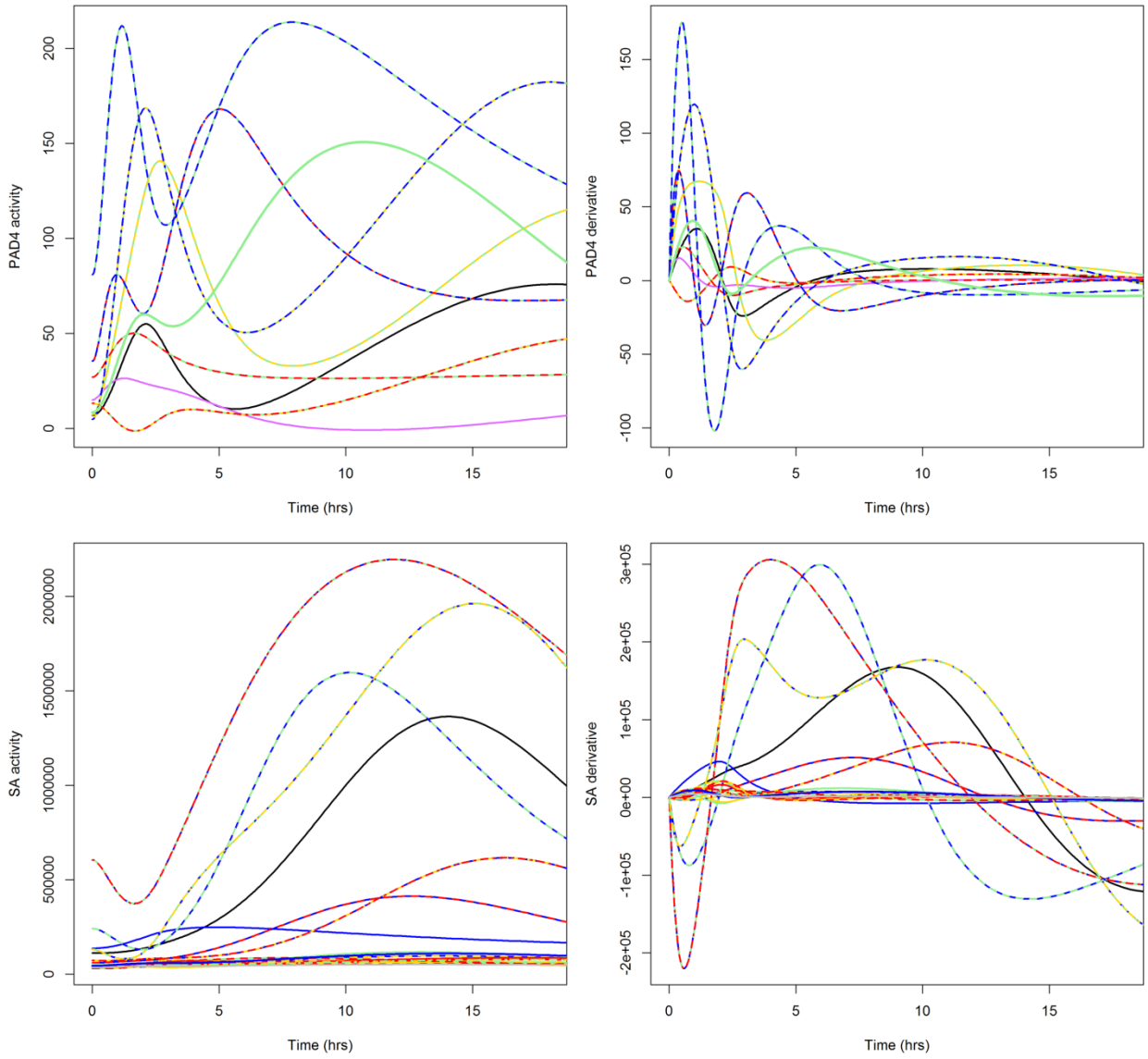


Figure 3-17. Time-interpolated PAD4 and SA sector activity data and their derivatives, in concentration units (y-axis). Left: Activity; Right: Derivatives; Upper: PAD4, Lower: SA. Black curve, wild type; purple curve, *fls2*. Other genotypes are denoted using colors that indicate sector presence. Red, JA; yellow, ET; green, PAD4; blue, SA. For example, the green and blue curve is the interpolated time course for *jePS*.

3.4.2 Delay Differential Equations (DDEs) can describe the triple mutant sector activities

As with the ND2D models in section 3.3, we used a stepwise discovery process for DDE-based models. The fundamental idea behind network reconstruction is that when deciphering regulatory mechanisms in a complex network, wipe out the network first and then build it back one component at a time. Following this strategy, we first sought models that could explain the triple mutant behavior of each sector. Note that this is a major improvement over the ND2D models, where the triple mutant activity was assumed, not explained.

With only a few assumptions, a single set of related DDE model structures could explain the triple mutant behavior for the JA, ET, and PAD4 sectors in *Jeps*, *jEps*, and *jePs*, respectively. We did not seek to explain the SA triple mutant behavior, since there are no significant changes in the SA concentrations in the *jepS* triple mutant. As seen in the Kim model (Kim et al. 2014), and in the work described in Chapter 4, SA is entirely dependent on the network for activation by flg22-dependent signaling, and is unresponsive on its own.

The activity values of all four sectors in most of the genotypes where that sector is active show two distinct peaks (see Figures 3-16 and 3-17; for the SA sector, the first peak is downwards). The width of the first peak is roughly one tenth of the width of the second peak. A reasonable modeling assumption for a signal transduction network is that an input pulse enters the network after an external stimulus. However, it is unlikely that a single input pulse into the network could create both of these peaks via feedback within the network. There are no simple feedback loops in the signal processing literature that we know of that can create such strong peak distortion, turning a sharp input pulse in a second very wide pulse; feedback can modulate peak width, but not to the extent seen in our data. To generate such a broad second peak with a simple mechanism creates a network that never returns to the steady state level after an input stimulus. A simpler hypothesis is that there are two different input pulses into the network, a fast, sharp input pulse, and a slow, wide input pulse. Biologically, it is entirely possible for there to be multiple signaling routes by which the input signal from flg22 recognition enters the

JA-ET-PAD4-SA network. Some of these routes could involve relatively few steps, or very fast signaling steps, and thus enter the network quickly, while other routes may involve time-consuming signal processing steps, thus widening and delaying the input signal. We note that in the Kim model (Kim et al. 2014), there were two independent input signal flows from flg22 to each of the sectors JA, ET, and PAD4. Our modeling assumption here of two independent inputs from flg22 to each of these sectors is consistent with the published Kim model (Kim et al. 2014).

For differential-equation-based systems to return to steady-state, a decay term is needed; otherwise the area under the input pulses is integrated and permanently offsets the long-term steady-state level of a sector. We therefore assumed a standard decay term proportional to sector activity in each triple mutant model, in addition to two external input pulses. What is a bit unusual in our models is an assumed delay in the sector auto-feedback decay term. These delays, unique to each sector, are necessary to produce the triple mutant sector activity shapes we observed. As described in section 3-3, delays in signal processing are a biologically reasonable feature for models of our reduced network.

The two input pulses were modeled as Beta-distributions since the Beta-distribution is a flexible peak shape that allows for different rising and falling speeds, unlike a Gaussian input pulse. For JA and ET, the following models were fit to the interpolated triple mutant concentration data (*Jeps* for JA and *jEps* for ET):

$$\frac{dJA}{dt} = first\ peak + second\ peak + k_{JA} \cdot JA_{k_{delay,JA}}$$

Equation 3-5.

$$\frac{dET}{dt} = first\ peak + second\ peak + k_{ET} \cdot ET_{k_{delay,ET}}$$

Equation 3-6.

Here,

$$first\ peak = fp_{amp} \cdot \beta_{fp_\alpha, fp_\beta}$$

Equation 3-7.

where fp_{amp} is an amplitude, and $\beta_{\alpha,\beta}$is a Beta distribution with the shape parameters $\alpha = fp_{\alpha}$ and $\beta = fp_{\beta}$:

$$\beta_{fp_{\alpha},fp_{\beta}} = \frac{1}{\mathrm{B}(fp_{\alpha},fp_{\beta})} \cdot x^{fp_{\alpha}-1} \cdot (1-x)^{fp_{\beta}-1}$$

Equation 3-8.

x is defined on the unit interval, $[0,1]$, and $\mathrm{B}(fp_{\alpha},fp_{\beta})$ is a normalization factor, the Beta function, which is also called the Euler Intergral of the first kind:

$$\mathrm{B}(fp_{\alpha},fp_{\beta}) = \int_0^1 y^{fp_{\alpha}-1} \cdot (1-y)^{fp_{\beta}-1} \cdot dy$$

Equation 3-9.

The parameters fp_{start} and fp_{stop} are used to scale the interval where the Beta distribution is defined, $[0,1]$, onto the time interval $[fp_{start},fp_{stop}]$. Similarly, the second input peak is also defined by an independent Beta distribution:

$$second\ peak = sp_{amp} \cdot \beta_{sp_{\alpha},sp_{\beta}}$$ **Equation 3-10.**

Which is scaled from $[0,1]$ onto the time interval $[sp_{start},sp_{stop}]$. The parameters fp_{start} and sp_{start} are effectively delay parameters for the two input pulses.

For JA and ET, the steady-state sector activity values are effectively zero, but there is a non-zero offset for PAD4. Thus, for the PAD4 model, a constant input $basal$ was added to the model to capture this observation:

$$\frac{dPAD4}{dt} = first\ peak + second\ peak + basal + k_{PAD4} \cdot PAD4_{k_{delay,PAD4}}$$

Equation 3-11.

For all three sectors, the sector auto-feedback is delayed, e.g., $PAD4_{k_{delay,PAD4}} = PAD4\ (t - k_{delay,PAD4})$. The decay parameter k_S where S is one of the

sectors JA, ET, or PAD4, quantifies the strength of the sector auto-feedback and $k_{delay,S}$ quantifies the delay in sector auto-feedback for each sector S. Note that for stable system behavior k_S must be a negative number.

Interpolated sector activity concentration values were scaled individually for each sector; the maximum observed sector activity in any genotype was set to 100. This was done to avoid the computational overhead when multiplying very large numbers, like those seen in the sector activity concentrations. DDE model fits to scaled, interpolated triple mutant data are shown for JA, ET, and PAD4 in Figures 3-18, 3-19, and 3-20, respectively. Equations were solved using the MATLAB solver `dde23` (Shampine and Thompson 2001). Since the models are non-linear, reasonably good model parameters were found by hand; fits were then optimized using the MATLAB optimizer `lsqcurvefit`. The optimized models fit the time-interpolated concentration data very well: Pearson correlation values of 0.998, 0.999, and 1.000, were seen for JA, ET, and PAD4 sector activities, respectively. It should be noted, however, that these models use more parameters than the number of underlying mean data points (7 time points). Thus, model fitting was made possible by generating more time points for the data through time-interpolation.

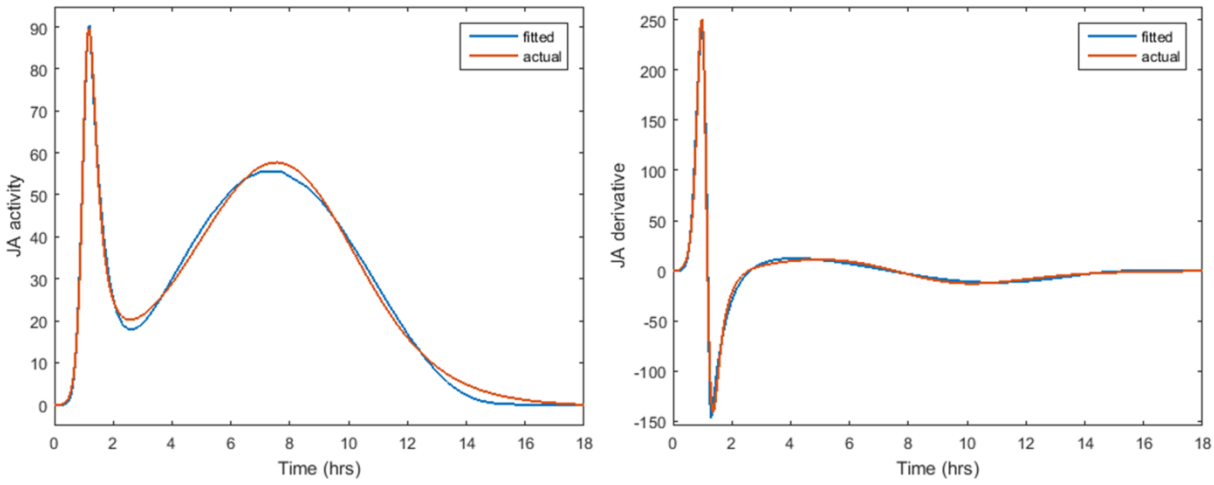


Figure 3-18. A DDE model fits time-interpolated JA activity values in the triple mutant *Jeps* very well. JA activity values (left) and JA derivative values (right) were fit to the model in Equation 3-5. Interpolated values are in orange; modeled values are in blue. The models were fit on t ∈ [0, 60] hrs; correlation was calculated for the full interpolated range t ∈ [0,36] hrs. Correlation between activity and modeled activity: 0.998.

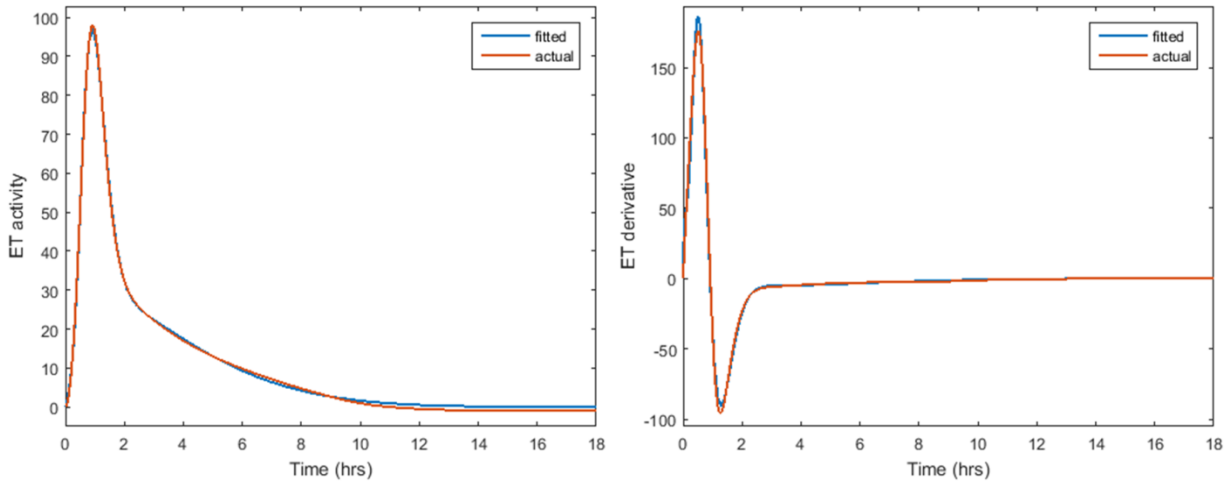


Figure 3-19. A DDE model fits time-interpolated ET activity values in the triple mutant *jEps* very well. ET activity values (left) and ET derivative values (right) were fit to the model in Equation 3-6. Interpolated values are in orange; modeled values are in blue. The models were fit on $t \in [0, 18]$ hrs; correlation was calculated for the full interpolated range $t \in [0,36]$ hrs. Correlation between activity and modeled activity: 0.999.

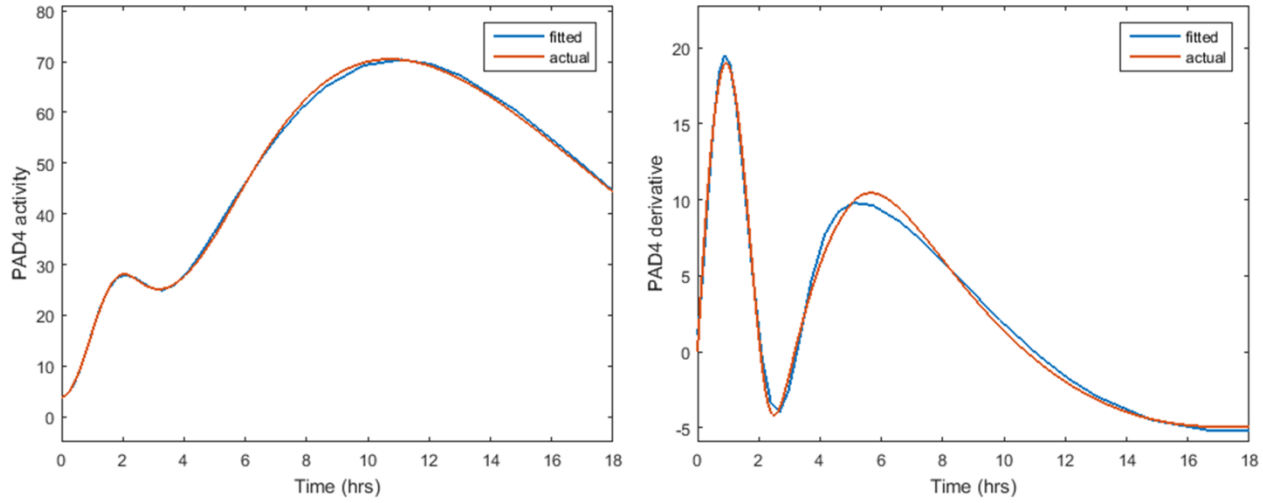


Figure 3-20. A DDE model fits time-interpolated PAD4 activity values in the triple mutant *jePs* very well. PAD4 activity values (left) and PAD4 derivative values (right) were fit to the model in Equation 3-11. Interpolated values are in orange; modeled values are in blue. The models were fit on $t \in [0,60]$ hrs; correlation was calculated for the full interpolated range $t \in [0,36]$ hrs. Correlation between activity and modeled activity: 1.000.

3.4.3 A DDE-based model of PAD4 sector activity

Relatively simple rules could explain the triple-mutant behavior for JA, ET and PAD4 (section 3.4.2). However, careful inspection of the concentration-based activity curves suggested that the ways that additional sectors modulate these triple-mutant responses is not simple. As one example among many, JA activity in the double mutant *JEps* is nearly identical to that of the triple mutant *Jeps* for the first peak, but considerably different for the second peak (Figure 3-16). A clear trend holds for all four sectors that the differences in sector activity between any genotype and the triple mutant appear to follow different trends in the first peak and second peak. Different rules for how the other sectors influence the two different input peaks is an entirely reasonable assumption: hundreds of different molecular factors influence and are influenced by the four major sectors of the plant immune signaling network; if even a few of these preferentially affect one peak, different rules could emerge for the two peaks. We therefore allowed separate mathematical terms to be used to describe these different effects on the two peaks. However, when allowing such a flexible DDE model, the number of parameters needed to describe any sector's activity across all eight genotypes where that sector is active could easily approach or exceed the number of mean data points underlying the data time-interpolation. This does not mean that such models would necessarily be incorrect, but any such models are likely fitting interpolation artifacts as well as network mechanisms.

Nonetheless, these models might make testable predictions, which can then be investigated. With this in mind, my advisor, Fumi Katagiri, manually discovered a DDE-based mathematical structure that appeared to capture the major features of the PAD4 sector activity time courses (Equation 3-12). I fit this model in MATLAB, using his starting parameters, and I optimized the parameters. A cartoon of this PAD4 model structure is seen in Figure 3-21. Note that the PAD4 sector activities from the complete stepwise network reconstitution genotype set are highly dynamic and that these dynamics are quite different between the different genotypes (Figure 3-16). These data strongly constrained the model discovery process, requiring the model to be biologically consistent across all network reconstitution states.

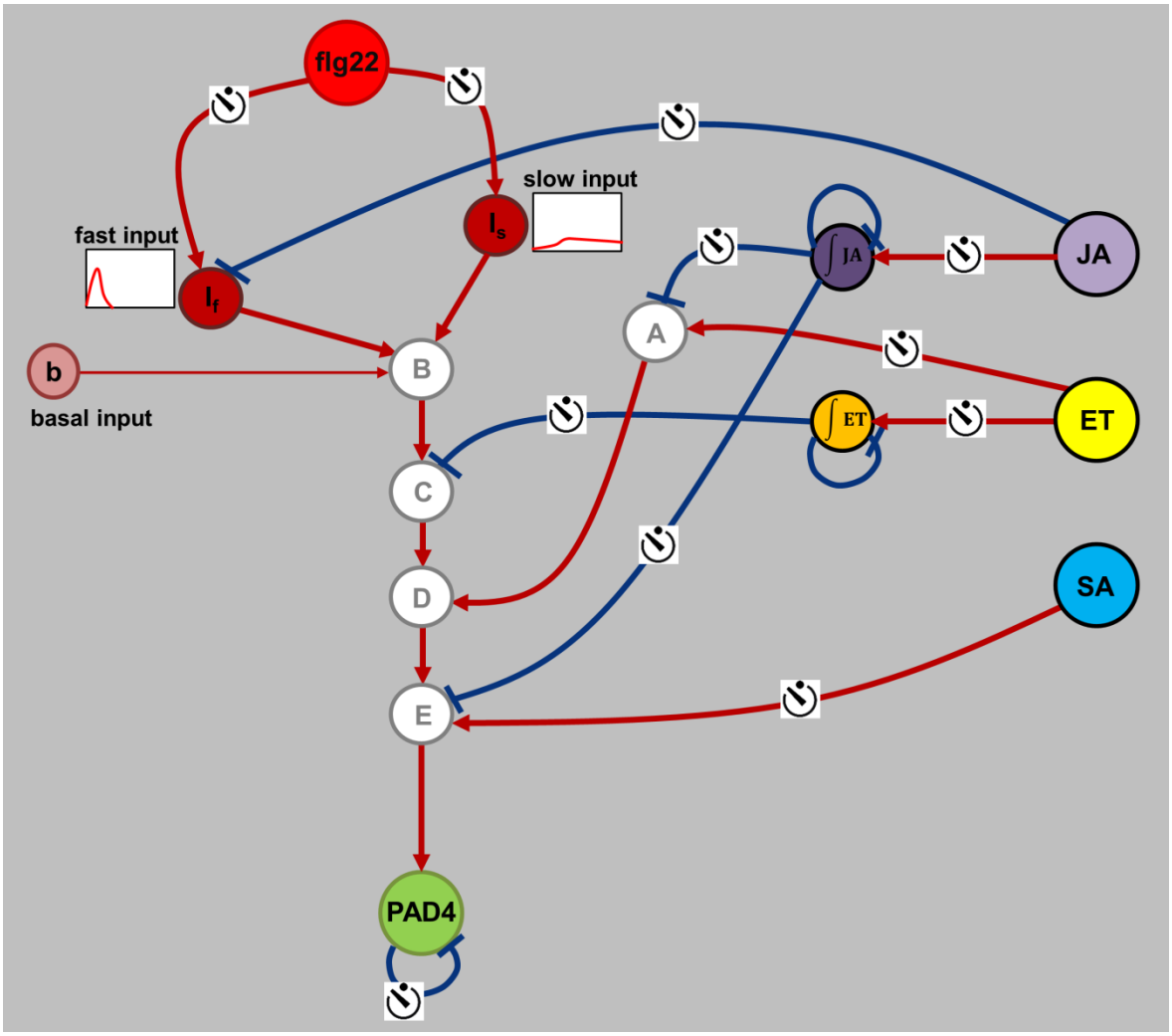


Figure 3-21. A diagram of our DDE-based model of PAD4 sector activity. As in the triple mutant, there are three inputs into the system: a fast input (I_f), which induces the first peak, a slow input (I_s), which induces the second peak, and a basal input, which allows non-zero steady-state behavior. Additionally, the PAD4 sector itself experiences auto-decay with a delay. The timer symbol indicates a delay in the given edge. The activities of the added sectors, JA, ET and SA, modulate the system inputs either directly, or via nodes which integrate those activities. The nodes A, B, C, D and E are predicted nodes where signals converge and are combined.

The PAD4 model (see below for mathematical details), has several key features. First, in addition to inputs directly from JA , ET, and PAD4 to the signals affecting PAD4, two integrating nodes, $\int JA$, which I call JA_{int} below, and $\int ET$, which I call ET_{int} below, integrate and thereby store transient information from JA and ET for longer periods of time. Second, flexible delays are present between node activities and their influences. Finally, the PAD4 model requires several intermediate nodes which combine converging

influences in different ways. These intermediate nodes (A, B, C, D, and E in Figure 3-21) are model predictions; if this model is correct, biological responses should exist which track the activity levels of these intermediate nodes. In the future, our RNA-Seq dataset is a natural place to search for proxies of these predicted nodes.

Mathematically, the full PAD4 model structure is:

$$\frac{dPAD4}{dt} = \Big(\big(first\ peak \cdot JA_{fp} + second\ peak + basal\big) \cdot ET_{int,C} + ET_A \cdot J_{int,A}\Big) \cdot SA_E \cdot JA_{int,E} + k \cdot PAD4\ (t - k_{delay})$$

Equation 3-12.

JA_{fp} is the JA influence on the first peak, assumed to take on a Michaelis-Menten-like inhibition:

$$JA_{fp}(t) = \frac{JA_{fp,asym}}{\left(1 + \left(\frac{JA_{fp,K_m}}{JA(t - JA_{fp,delay})}\right)^{JA_{fp,Hill}}\right)}$$

Equation 3-13.

Where $JA_{fp,asym}$, JA_{fp,K_m}, and $JA_{fp,Hill}$ are the asymptote, K_m and Hill coefficients of a Michaelis-Menten relationship. $JA(t - JA_{fp,delay})$ is the delayed JA activity, delayed by the time amount $JA_{fp,delay}$. Assuming Michaelis-Menten relationships allows influences to saturate, a common assumption in modeling biochemical processes.

The ET_A and SA_E terms in Equation 3-12 are the ET influence on the A node, and the SA influence on the E node, respectively. Both of these influences are also modeled using Michaelis-Menten relationships, but assuming activation, rather than inhibition:

$$ET_A(t) = 1 + \frac{ET_{A,asym}}{\left(1 + \left(\frac{ET_{A,K_m}}{ET(t - ET_{A,delay})}\right)^{ET_{A,Hill}}\right)}$$

Equation 3-14.

SA_E is defined similarly, with the coefficients $SA_{E,asym}$, SA_{E,K_m}, $SA_{E,delay}$, and $SA_{E,Hill}$. The integrated nodes, ET_{int}, and JA_{int}, which integrate ET and JA are calculated by solving the ordinary differential equation

$$\frac{dET_{int}(t)}{dt} = ET\left(t - ET_{delay}\right) + k_{ET}ET_{int}(t)$$

Equation 3-15.

Where ET_{delay} and k_{ET} are constants; JA_{int} is defined analogously. The terms $ET_{int,C}$, $JA_{int,A}$, and $JA_{int,E}$ in Equation 3-12 are Michaelis-Menten inhibitions of the form in Equation 3-13, with four associated parameters each: $asym$, K_m, $delay$, and $Hill$.
Note that terms that involve JA and ET are only defined when those sectors are present. Terms involving SA, however, are defined in all PAD4-containing genotypes, since basal SA and slight variations in these basal levels are present, even in *sid2*-containing genotypes.

The starting parameters for first and second input peaks, the basal input, and the PAD4 auto-feedback delay and decay parameters were first optimized using the triple mutant *jePs* PAD4 time-interpolated sector activity curve. Then the PAD4 sector activity curves in all the PAD4-containing genotypes, *JEPS*, *jEPS*, *JePS*, *JEPs*, *jePS*, *jEPs*, *JePs*, and *jePs*, were fit simultaneously.

By exploring model behavior by hand using an integrator he built in R, my adviser (Katagiri, unpublished) discovered starting values for all 41 parameters (see Table 3-4). I used his values as starting parameters, and optimized them using MATLAB's `lsqcurvefit` optimizer wrapped around the MATLAB DDE solver `dde23` (Shampine and Thompson 2001). Node values for JA_{int} and ET_{int} were calculated using the ordinary (not delay) ODE solver `ode45`.
The upper and lower bounds for parameters were set arbitrarily but tightly around the initial parameter values, holding parameters from straying by more than order of magnitude (see Table 3-4). Four additional optimizations were then performed; for the first two, the parameter bounds were loosened 10% at a time; the model parameters did

not change much during the second of these two, which was the third optimization in total. For the next two optimizations, parameter bounds were loosened by an order of magnitude each time; again, the model parameters changed little during the second iteration. The result was a model that fit the interpolated data well (Figure 3-22). Pearson correlation between the interpolated PAD4 activity values and the modeled values was 0.944, indicating excellent agreement between the model and the interpolated data. The agreement between data and model go beyond just average correlation, however. The timing of nearly all the peaks and valleys were accurately captured by the optimized model. The model also provides an interesting explanation for a notable dynamic feature in the PAD4 data: when the ET sector is present, it appears to delay first PAD4 peak (Figure 3-17). According to the PAD4 model, this is caused by ET shutting off the inputs to PAD4, and instead regulating PAD4 itself, together with an integrated signal from JA. The model fits well, but at the cost of numerous parameters: 41 in total. The time-interpolated data values are only supported by 56 mean values (7 time points · 8 genotypes = 56). To support a model with this many parameters, we assumed that our time-interpolated data are reliable.

Predicted intermediate nodes A, B, C, D, and E (see Figure 3-21) were identified as follows: Whenever multiple converging influences in Equation 3-12 alternated between additive and multiplicative terms, an intermediate node was inserted. As an example, the three additive terms in the expression $first\ peak \cdot JA_{fp} + second\ peak + basal$ converge onto node B. These terms are all multiplied by $ET_{int,C}$, therefore this integration occurs at the next predicted node, C.

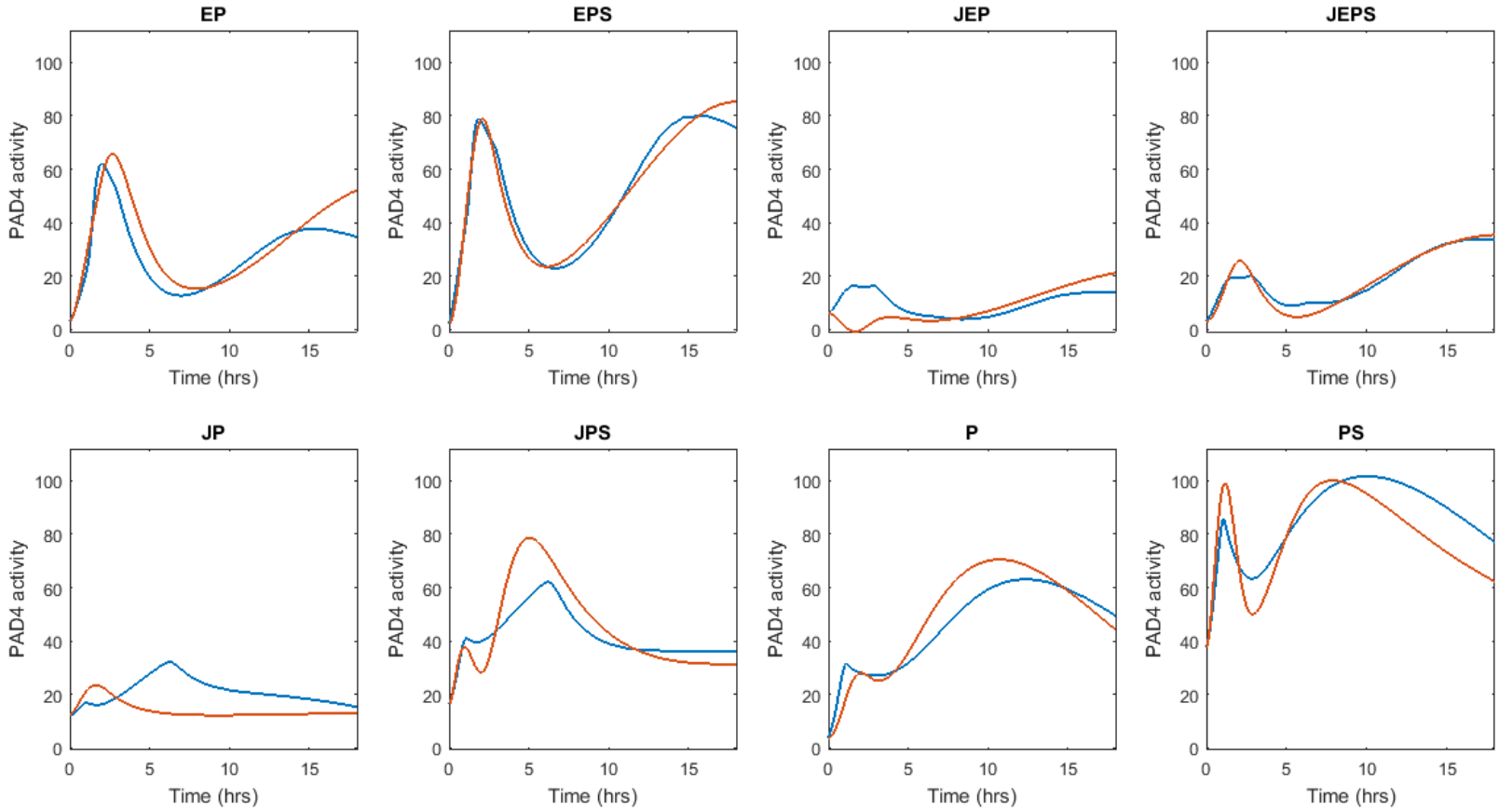

EP
EPS
JEP
JEPS
JP
JPS
P
PS
PAD4 activity
Time (hrs)
0
20
40
60
80
100
0
5
10
15

Figure 3-22. (See previous page.) **A DDE model fits time-interpolated PAD4 activity values in all PAD4-containing genotypes well.** PAD4 activity and derivative values for the genotypes *JEPS*, *jEPS*, *JePS*, *JEPs*, *jePS*, *jEPs*, *JePs*, and *jePs* were simultaneously fit to the model in Equation 3-12. For simplicity, the panel for each genotype is labeled by the active sectors only, for example *jePs* is P. Interpolated values are in orange; modeled values are in blue. The models were fit on $t \in [0,60]$ hrs; correlation was calculated for the full interpolated time range $t \in [0,36]$ hrs. Correlation between activity and modeled activity: 0.944.

Parameter description	Final optimized values	Initial optimized values	Initial starting values	Initial lower bounds	Initial upper bounds	Final starting values	Final lower bounds	Final upper bounds
fast input start	0.46	0.48	0.16	0	0.5	0.46	0	0.5
fast input stop	1.55	2.28	4.43	1	3.5	1.55	1	60
fast input amplitude	7.96	6.34	14.66	5	12	7.96	0	1000000
fast input alpha	2.56	3.18	3.28	3	10	2.55	1	1000
fast input peak time	1.22	2.05	1.44	0.5	4.43	1.22	0.1	18
slow input start	0	0.42	0.02	0	5	0	0	5
slow input stop	59.87	24.11	60	18	60	59.87	1	60
slow input amplitude	1.38	8.01	24.89	8	25	1.38	0	1000000
slow input alpha	1.82	1.51	3.32	1	18	1.82	1	1000
slow input peak time	7.86	15.78	10.37	5	18	7.86	0.1	18
PAD4 auto-feedback strength	-0.61	-2	1	-2	-0.5	-0.61	-20	20
PAD4 auto-feedback delay	0	0.01	0.1	0	1	0	0	100
basal input	0.74	1	1.1	1	1.8	0.74	-100	100
SA influence on E node; asymptote	16.96	5	11.5	5	15	16.96	0.04	1815
SA influence on E node; Km	3.66	10	7	5	10	3.66	0.04	1210
SA influence on E node; hill coefficient	1.54	1	1.8	1	2	1.54	0.01	242
SA influence on E node; delay	0	0.01	0.2	0	0.5	0	0	60.5
Intergrated JA influence on E node; asymptote	1.38	1	1	0.5	2	1.38	0	242

Intergrated JA influence on E node; Km	10.22	35.41	38	30	45	10.22	0.24	5445
Intergrated JA influence on E node; hill coefficient	-0.39	-1	-1.5	-2	-1	-0.39	-242	-0.01
Intergrated JA influence on E node; delay	5.7	3.97	3.7	2.5	4	5.7	0.02	484
JA influence on fast input; asymptote	0.59	0.59	1	0.5	1.5	0.59	0	181.5
JA influence on fast input; Km	2.43	30	35	30	40	2.43	0.24	4840
JA influence on fast input; hill coefficient	-0.26	-2.5	-1.8	-2.5	-1	-0.27	-302.5	-0.01
JA influence on fast input; delay	0.5	0.43	0.1	0	0.5	0.5	0	60.5
Intergrated ET influence on C node; asymptote	0.5	0.5	1	0.5	2	0.5	0	242
Intergrated ET influence on C node; Km	9.67	7	5	3	7	9.66	0.02	847
Intergrated ET influence on C node; hill coefficient	-1.96	-0.57	-1	-1.5	-0.5	-1.96	-181.5	0
Intergrated ET influence on C node; delay	0.1	0.1	0.1	0	0.4	0.1	0	48.4
Intergrated JA influence on A node; asymptote	0.08	0.5	1	0.5	1.5	0.08	0	181.5
Intergrated JA influence on A node; Km	1.79	22	30	22	37	1.79	0.18	4477
Intergrated JA influence on A node; hill coefficient	-1.2	-0.5	-1	-1.5	-0.5	-1.2	-181.5	0
Intergrated JA influence on A node; delay	0.1	0.1	0.1	0	0.4	0.1	0	48.4
ET influence on A node; asymptote	783.8	230.06	300	230	350	783.23	1.86	42350
ET influence on A node; Km	166.59	189.98	150	120	190	166.58	0.97	22990
ET influence on A node;	3.12	3.7	3	2.3	3.7	3.12	0.02	447.7

hill coefficient								
ET influence on A node; delay	0.5	0.5	0.5	0.1	1	0.5	0	121
Integrated JA node auto-feedback strength	0.08	1	2	1	3	0.08	0.01	363
Integrated JA node auto-feedback delay	-0.44	-0.12	-0.25	-0.5	-0.1	-0.44	-60.5	0
Integrated ET node auto-feedback strength	1.93	2.5	1.5	1	2.5	1.93	0.01	302.5
Integrated ET node auto-feedback delay	-1.22	-0.55	-0.3	-0.55	-0.1	-1.22	-66.55	0

Table 3-4. Parameter values for the PAD4 DDE model optimization. Final optimized values are the parameter values after all 5 optimization runs; initial optimized values are the parameter values after the first model optimization.

3.5 Future Directions

Despite having an order of magnitude more data than other flg22 profiling experiments to date, it became apparent through the modeling efforts detailed in this chapter that more data collected strategically at important time points and in critical genotypes still are required for reliable mechanistic models of the JA-ET-PAD4-SA network. The models discussed in this chapter rely on time-interpolated data; for reliable models we need more confidence in the shape of the sector activity time courses. The network dynamics are more complex than our current data: simple models don't fit well; complex models fit the time-interpolated data well, but they risk fitting time-interpolation artifacts. A clear priority for future modeling efforts is to strategically collect additional data to more confidently determine the shape of the sector activity time-interpolation curves, and thereby increase our confidence that the models fit to these curves capture the mechanisms underlying the network, and not merely time-interpolation artifacts. Specifically, more detail will be required where our current sector activity interpolations are highly dynamic.

It can be a major experimental challenge to simultaneously collect infiltrated plant leaf tissue from 16 different genotypes. It may be worthwhile to collect time courses with additional detail at strategic time points for subsets of these 16 genotypes, and to produce models that capture portions of the network. For example, two triple mutants and a double mutant could be used to model each two-sector sub-network. Alternatively, balanced experimental designs could be used to collect all needed data from all genotypes via several different tissue collection experiments with overlapping subsets of genotypes. This data could then all be combined statistically using a mixed-effects linear model.

The PAD4 DDE model was fit using the best of several genes that passed our criteria for a PAD4 sector activity proxy (see Chapter 4 for a list of alternative sector activity proxies for PAD4). To increase confidence in the edges and nodes of the PAD4 model, models should be discovered and fit to other candidate genes as well.

The DDE-based models show the greatest promise for having the needed flexibility to capture dynamic sector behavior, and should be used for future modeling efforts. However, the current form of the PAD4 DDE model is lacking one critical feature: the sector activity inputs from the other sectors are from measured data, not model-predicted data using DDE models of those other sectors. A complete DDE-based network model needs to connect all four sectors via 4 coupled DDE equations, so that given only input information, network dynamics across all four sectors, and throughout the time course, can be accurately predicted. This type of free-running model is the standard for mechanistic modeling. Such models are able to generate important type II predictions: how the network will behave under different input conditions.

It may also be worth developing a complementary experimental platform for monitoring sector activity. Given the complexity in the data presented in this chapter, it is my opinion that numerous additional strategically chosen time points will be critical to producing highly reliable time-interpolations. The RNA-Seq data allowed me to find reliable gene expression proxies for PAD4 and ET, and undoubtably RNA-Seq data will prove useful in mining for predicted node proxies, but the deep sampling in gene space provided by RNA-Seq may not be necessary for the next round modeling efforts. Rather, deep time

sampling is the priority. It may be possible to simultaneously monitor fluorescent or bioluminescent reporters hooked up to the promoters of selected proxies for each sector. Different reporters and separate channels could be used to simultaneously monitor several sector activities. This kind of data might provide not only the extensive time sampling likely needed to construct reliable DDE models for the JA-ET-PAD4-SA network, but would also provide extended time-series measurements from individual tissue samples, rather than relying on different plants for different time samples, a limitation of the current data.

Chapter 4.

The Highly Buffered Arabidopsis Immune Signaling Network Hides the True Function of its Components

The material in this chapter has been submitted for publication to the journal PLOS Genetics. I performed the data processing and spearheaded the data analysis in this chapter, in collaboration with my adviser, Fumi Katagiri. Chad Myers at the University of Minnesota provided feedback on conceptual content. The data used in this chapter were generated from flg22 infiltration experiments performed by Kenichi Tsuda, who was at the University of Minnesota at the time. Tag-Seq libraries (Rallapalli et al. 2014) for expression profiling generated from these plant tissue samples were mainly made by Kenichi Tsuda, Ghanashyam Rallapalli, and Shuta Asai, with some assistance from William Truman and myself; Ghanashyam's and Shuta's work was done in the laboratory of Jonathan Jones at The Sainsbury Laboratory in Norwich, UK; William's work was done at the University of Minnesota. Hormone profiles were generated by Hitoshi Sakakibara at the RIKEN Center for Sustainable Resource Science in Japan, with assistance from Mikiko Kojima. Motif analysis work was performed by Matthew Papke, an undergraduate in Computer Science at the University of Minnesota, under my supervision. I wrote the paper, together with my adviser, Fumi Katagiri.

Authors

Rachel A. Hillmer[1], Kenichi Tsuda[1, 2], Ghanasyam Rallapalli[3], Shuta Asai[3, a], William Truman[1, b], Matthew D. Papke[1, 4], Hitoshi Sakakibara[5, c], Jonathan D. G. Jones[3], Chad L. Myers[4], Fumiaki Katagiri[1]

Author Affiliations

[1] Department of Plant Biology, Microbial and Plant Genomics Institute, University of Minnesota, Twin-Cities, Saint Paul, MN 55108, USA
[2] Department of Plant-Microbe Interactions, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany
[3] The Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, UK

[4] Department of Computer Science and Engineering, University of Minnesota, Twin-Cities, Minneapolis, MN 55455, USA
[5] RIKEN Center for Sustainable Resource Science, Yokohama 230-0045, Japan
[a] Current Address: Center for Sustainable Resource Science, RIKEN, 1-7-22 Suehiro-cho, Tsurumi, Yokohama, Kanagawa 230-0045 Japan
[b] Current Address: Institute of Plant Genetics of the Polish Academy of Sciences, Poznan, Poland
[c] Current Address: Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8601, Japan

4.1 Summary

Plant immunity protects plants from numerous potentially pathogenic microbes. The biological network that controls plant inducible immunity must function effectively even when network components are targeted and disabled by pathogen effectors. Network buffering could confer this robustness by allowing different parts of the network to compensate for loss of each other's functions. Networks rich in buffering rely on interactions within the network, but these mechanisms are difficult to study by simple genetic means. Through a network reconstitution strategy, where we disassemble and stepwise reassemble the plant immune network that mediates Pattern-Triggered-Immunity, we have resolved systems-level regulatory mechanisms underlying the Arabidopsis transcriptome response to the immune stimulant flagellin-22 (flg22). These mechanisms show widespread evidence of interactions among major sub-networks—we call these sectors—in the flg22-responsive transcriptome. Many of these interactions result in network buffering. Resolved regulatory mechanisms show unexpected patterns for how the jasmonate (JA), ethylene (ET), phytoalexin-deficient 4 (PAD4), and salicylate (SA) signaling sectors control the transcriptional response to flg22. We demonstrate that many of the regulatory mechanisms we resolved are hidden from the traditional genetic approach of single-gene null-mutant analysis. As potential pathogenic perturbations to the network, null-mutant effects on the immune signaling network can be buffered by the network as well.

4.2 Introduction

A major tenet of systems biology is that complex biological systems are more than the sum of their parts (Kitano 2002). This literally means that interactions among system components are important. Because of these interactions, the function of a single gene in a complex signaling network may not be reliably inferred from null-mutant analysis of the gene alone; its function may be buffered by those of gene(s) in some other part of the network. To correctly assign functions to system components, the interactions among these components must be quantified.

Buffering has two different biological sources (Rutherford 2000; Hartman, Garvik, and Hartwell 2001): (I) closely related gene family members with similar function may buffer each other, and (II) unrelated biological functions may buffer each other's contribution to overall network function. Here we focus on the second case, which we call network buffering. Network buffering has been explored systematically in yeast, where the contributions of genetic interactions to yeast growth have been documented on a genome-wide scale (Costanzo et al. 2010).

Plants are faced with a barrage of pathogen assaults during their lifetimes. Plants deter pathogen attacks using both preformed barriers (Nawrot et al. 2014; Underwood 2012) and an inducible immune system; the latter resembles innate immunity in animals (Dodds and Rathjen 2010). In principle, microbial pathogens can evolve much faster than their hosts, and thus the hosts appear to face a serious adaptive disadvantage. In plants, which lack adaptive immunity, a highly buffered immune signaling network could confer robustness against perturbations to network components, as pathogen effectors are known to target plant immune signaling proteins (Asai and Shirasu 2015). Buffering conceals the identity of network components: when a component in a buffered network is disabled, the network can compensate to some degree for this loss.

The robustness of the plant immune signaling network is qualitatively different than the robustness of switch-like networks that control development and the metabolic responses to carbon source changes. These other networks ensure high-fidelity execution of an internal program despite internal and external noise. The plant immune

signaling network, in contrast, must deploy the right response intensity depending on attack severity—a fundamentally quantitative response, rather than a switch-like one, as immune responses carry a fitness cost for the plant (Vos, Pieterse, and Wees 2013). The plant immune signaling network needs to buffer not only random environmental noise, but more so, targeted component loss by pathogen effectors. By concealing the immune network's mechanisms from pathogens, network buffering may confer resilience to the plant against pathogen adaptation.

Network buffering conceals a network's underlying mechanisms not only from potential pathogens but also from our attempts to study the network. For example, null mutant analysis of individual genes in such a network may result in incorrect mechanistic interpretations (Tsuda and Katagiri 2010). How do we correctly assign component function in the presence of network buffering? One approach is to fully remove a given network function through deep network perturbation, and then stepwise reconstitute the network using combinatorial perturbations (Tsuda et al. 2009; Jansen 2003). To avoid a "combinatorial explosion", the network must contain only a small number of components. For large complex networks, like the network that mediates plant inducible immunity, one strategy is to reduce the network size using a more summarized scale.

A broad-spectrum form of inducible immunity, Pattern-Triggered-Immunity (PTI) can be stimulated by treating a plant with conserved microbe-associated molecular patterns (MAMPs), such as a portion of bacterial flagellin (flg22) (Bigeard, Colcombet, and Hirt 2015). The network that mediates flg22-PTI in Arabidopsis is almost entirely controlled by a network composed of four component sub-networks: the signaling mediated the plant hormones salicylate (SA), jasmonate (JA) and ethylene (ET) and by the major immune regulator, phytoalexin-deficient 4 (PAD4) (Tsuda et al. 2009). We call these sub-networks signaling sectors. Each of these sectors can be interrupted by disabling a single biosynthetic or signaling gene: *dde2-2* removes the JA sector (von Malek et al. 2002), *ein2-1* the ET sector (Alonso et al. 1999), *pad4-1* the PAD4 sector (Jirage et al. 1999), and *sid2-2* the SA sector (Wildermuth et al. 2001). We previously removed the JA-ET-PAD4-SA network *in planta* via a quadruple mutant *dde2 ein2 pad4 sid2* and stepwise reconstituted the network via 4 triple mutants, 6 double mutants, 4 single-gene mutants, and the wild type (Tsuda et al. 2009; Kim et al. 2014).

Here we collected transcriptome and hormone profiles of these combinatorial mutants along a detailed timecourse and quantified how interactions between the four major sectors JA, ET, PAD4 and SA regulate the dynamic plant transcriptome in response to flg22 treatment. We show that sector interactions, rather than individual functions, dominated the regulation of the transcriptome response, and that buffering was extensive in the network. We further demonstrated several specific cases where the functional mechanisms inferred from shallow network perturbation did not reflect the true regulation of the flg22-responsive and JA-ET-PAD4-SA network-dependent transcriptome. Our analysis spotlights a complex biological network that resists a classical genetics approach, a network where single-gene mutant analysis sometimes leads to incomplete and at points incorrect descriptions of component functions. We demonstrate the effectiveness of a systems biology approach, network reconstitution, to elucidate dynamical component functions of the plant immune signaling network.

4.3 Results and Discussion

4.3.1 Overview of data collection and analysis

We profiled leaves of 31-32 days old plants of 17 Arabidopsis genotypes 0, 1, 2, 3, 5, 9, and 18 hrs after flg22 treatment. Sample infiltration was staggered (S3 Text), so that samples were collected as close to the same time of day as was experimentally feasible, because of known influences of the circadian cycle on plant immunity. The genotypes used were our 16 combinatorial mutants for the 4 sectors, and a null mutant for *FLS2*, the flg22 receptor gene. In each mutant at each time point we profiled both the plant transcriptome via stranded mRNA Tag-Seq libraries (Rallapalli et al. 2014), and a suite of 44 plant hormones and related compounds via liquid chromatography-mass spectrometry (LC-MS) (Kojima et al. 2009; Kojima and Sakakibara 2012), including free SA and free JA. What we call transcript expression levels in this paper are approximately $\log_2$-transformed expression values; the transform was moderated at low expression levels (see Materials and Methods and S1 Text for details).

We selected genes and hormones for their dynamics and the sector effects on them. We focused in this paper on changes in transcript expression levels and did not consider network effects on basal transcript expression. For at least one genotype and at least one time point, the transcript expression (or abundance for hormones) level response (relative to 0 hrs) was required to be significantly different than the transcript response in the quadruple mutant. This value is referred to hereafter as the transcript response change of a gene transcript or hormone. The only active hormones that passed these criteria were free SA and free JA; 3211 genes passed these criteria, which hereafter we call the flg22-responsive and network-dependent genes. There were roughly as many flg22-responsive and network-independent genes in our dataset than there were flg22-responsive and network-dependent genes (S3 Fig). Nevertheless, focusing on flg22-responsive and network-dependent genes is appropriate since the JA-ET-PAD4-SA network is almost entirely responsible for flg22-triggered resistance (Tsuda et al. 2009).

Each of the selected 3211 genes and 2 hormones was individually subjected to signaling allocation analysis to estimate both the sector effects and sector interactions that causally control the gene or hormone at each time point; these estimated effects and interactions are called signaling allocations (Tsuda et al. 2009; Mateos et al. 2015).

4.3.2 The vast majority of the flg22-responsive and network-dependent genes showed complex network regulation

Of the 3211 genes and 2 hormones that showed a significant transcript (or hormone) response change, the signaling allocations of these response changes showed widespread evidence of complex network regulation (Figure 4-1). Signaling allocations are the quantitative contributions to relative responsive expression of the individual sectors (JA, ET, PAD4, SA), and their interactions, which are indicated using single-letter abbreviations and ‘:’ (J:E, J:P, J:S, E:P, E:S, P:S, J:E:P, J:E:S, J:P:S, E:P:S, and J:E:P:S) (see S1 Text for mathematical details about fitting signaling allocation models to gene expression data). When two individual contributions have the same sign, and the interaction between them also has the same sign, the two sectors synergize the

transcript response change. On the other hand, if two individual contributions have the same sign and the interaction between them has the opposite sign, the two sectors buffer each other (this is sometimes referred to as compensation). There is no simple description of a particular interaction (such as synergy or buffering) when the individual contribution signs are different. Nevertheless, the signaling allocations provide a complete quantitative description for how the JA, ET, PAD4 and SA sectors regulate the transcript response change of each gene or hormone. As reported for pathogen growth assays (Tsuda et al. 2009), the individual contributions of each sector to the transcript response changes in our dataset tended to all have the same sign, that is, all four positive, or all four negative (Figure 4-1; all red or all blue for the individual contributions). Given numerous literature cases that document antagonism between many of these sectors (JA and SA, JA and ET, SA and ET (Pieterse et al. 2012; Zheng et al. 2012; Van der Does et al. 2013; Broekgaarden et al. 2015; Chen et al. 2009)), our analysis suggests that the source of these antagonisms is likely mediated by higher-order interactions for most of the flg22-responsive and network-dependent genes.

The bulk of the flg22-responsive and network-dependent transcriptome critically required sector interactions to explain transcript behavior, including three- and four-sector interactions. 99% of the 3211 genes had at least one statistically significant interaction allocation. As a whole, the transcriptome response was skewed towards complex regulation (Figure 4-2): 71% of flg22-responsive and network-dependent genes showed more complex regulation than simple (single sector) regulation.

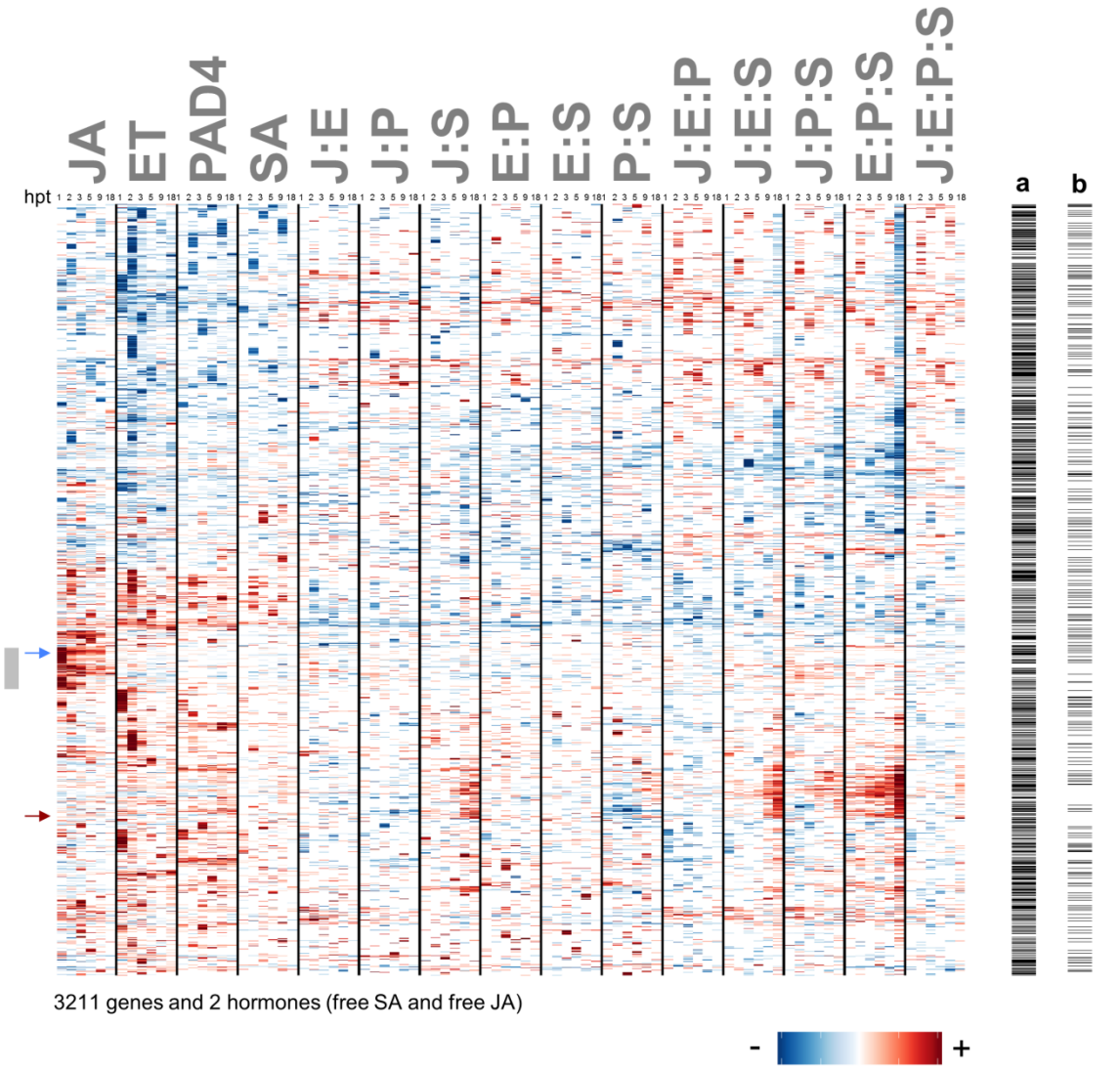


Figure 4-1. Network sector interactions are common in the regulation of the transcriptome response to flg22 treatment. Note that even three- and four-sector interactions are common. A clustergram is shown for hierarchical clustering of 3211 flg22-responsive network-dependent genes and the two hormones SA and JA (rows); clustering is based on their signaling allocations (columns). The sign and magnitude of the signaling allocation contributions are indicated by color and intensity; insignificant values are set to zero—visually they are white. Allocations were calculated for each gene or hormone from the transcript expression responses (abundance responses for hormones) through the time-series 1, 2, 3, 5, 9, and 18 hours post treatment (hpt); 6 minor columns representing the 6 time points post treatment follow each other consecutively within each major column. Individual sector contributions are shown in the first four major columns and are indicated as JA, ET, PAD4, and SA. Sector interactions are shown in the rest of the major columns and indicated using ':' with single letter notations of the sectors, for example the interaction between JA and ET is denoted 'J:E'. Columns a and b are indicator columns that mark, with black rows, (a) the 1691 genes whose transcript response relative to wild type is buffered in all single-sector mutants (*dde2*, *ein2*, *pad4*, and *sid2*), and (b) the 527 genes whose transcript response relative to wild type is buffered in all single and double-gene sector mutants (*dde2*, *ein2*, *pad4*, *sid2*, *dde2 ein2*, *dde2 pad4*, *dde2 sid2*, *ein2 pad4*, *ein2 sid2*, and *pad4 sid2*).The gray sidebar on the left of the clustergram marks a set

of genes that respond mainly to the JA sector. Blue and red arrows on the left of the clustergram indicate the rows corresponding to free JA and free SA, respectively. The 3211 flg22-responsive, network-dependent genes were selected according to the procedure described in the Materials and Methods. Only signaling allocation contributions with q-values < 0.05 (Storey's FDR) are shown. Signaling allocation rows were scaled for better visualization using the extreme value of each row.

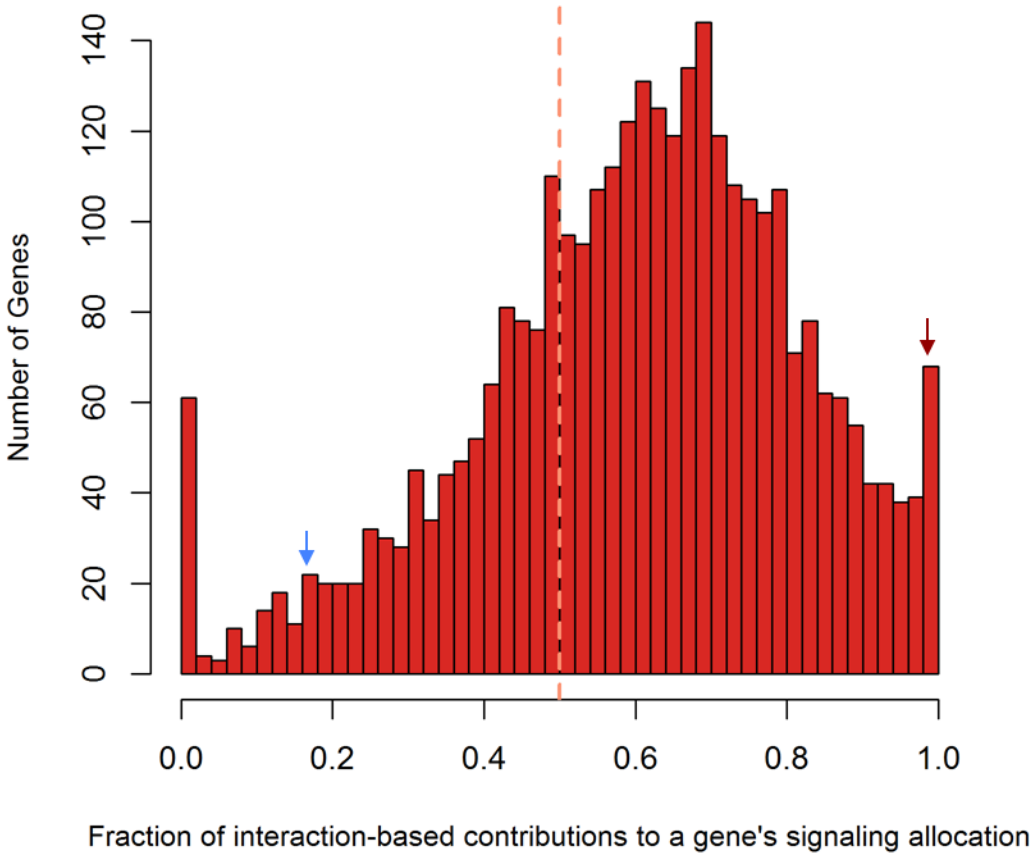


Figure 4-2. Network sector interactions dominate over individual single sector contributions to the flg22 transcriptome response. The distribution of how much sector interactions contribute to each transcript or hormone's signaling allocation is shown as a histogram, calculated from signaling allocations of the 3211 flg22-responsive, network-dependent genes and the hormones free SA and free JA shown in Figure 4-1. Genes whose transcript response changes were dependent only on additive effects of individual sector contributions have no complex contributions and thus a value of 0 for the fraction of interaction-based contributions to a gene's (or hormone's) signaling allocation. On the other end of the spectrum, genes whose transcript response changes were dependent only on interaction terms without any individual sector contribution have a value of 1 for the fraction of interaction-based contributions to a gene's (or hormone's) signaling allocation. Fractions were calculated via the Euclidean length of a vector representing only the sector interaction terms divided by the Euclidean length of the vector representing all signal allocation terms. Arrows indicate the bins containing the calculated fraction for JA (blue) and SA (red). 71% of the genes in the distribution have a value past 0.5, where complex and simple terms contribute to a signaling allocation with equal magnitude.

We have previously demonstrated that regulation of both a phenotypic output of this network (reduction in pathogen growth), as well as core network activity (the regulation of four sector activity markers, each for one of the JA, ET, PAD4, and SA sectors), is complex (Tsuda et al. 2009; Kim et al. 2014). Here we expand that observation, noting that thousands of flg22-responsive, network-dependent genes were causally driven by interactions within the JA-ET-PAD4-SA network. Interactions among network components are an essential feature of any network that exhibits strong network buffering.

4.3.3 Half of the flg22-responsive and network-dependent genes are well buffered.

Among the 3211 flg22-responsive and network-dependent genes, 1691 genes were fully buffered—they showed no significant transcript expression level change relative to the wild type response in any single-sector mutant (black rows in Figure 4-1 column a indicate these 1691 genes; buffering summary in Figure 4-3A); deeper network perturbations were required to show disturbances in network response. Furthermore, 527 genes among the 1691 genes showed no significant transcript change in any double mutants either (black rows in Figure 4-1 column b indicate these 527 genes; summary in Figure 4-3B). Yet these genes with clear network buffering had rich signaling allocation signatures. Thus, despite their lack of responses in single- and double-sector mutant genotypes, these genes may be key players in plant immunity because they were clearly regulated by the network, each responded strongly to an immune stimulus (flg22), and these responses were highly buffered from potential pathogenic perturbations: they remained like those of wild type despite strong network perturbation. S5 Text and S6 Text contain the AGI locus identifiers for these genes.

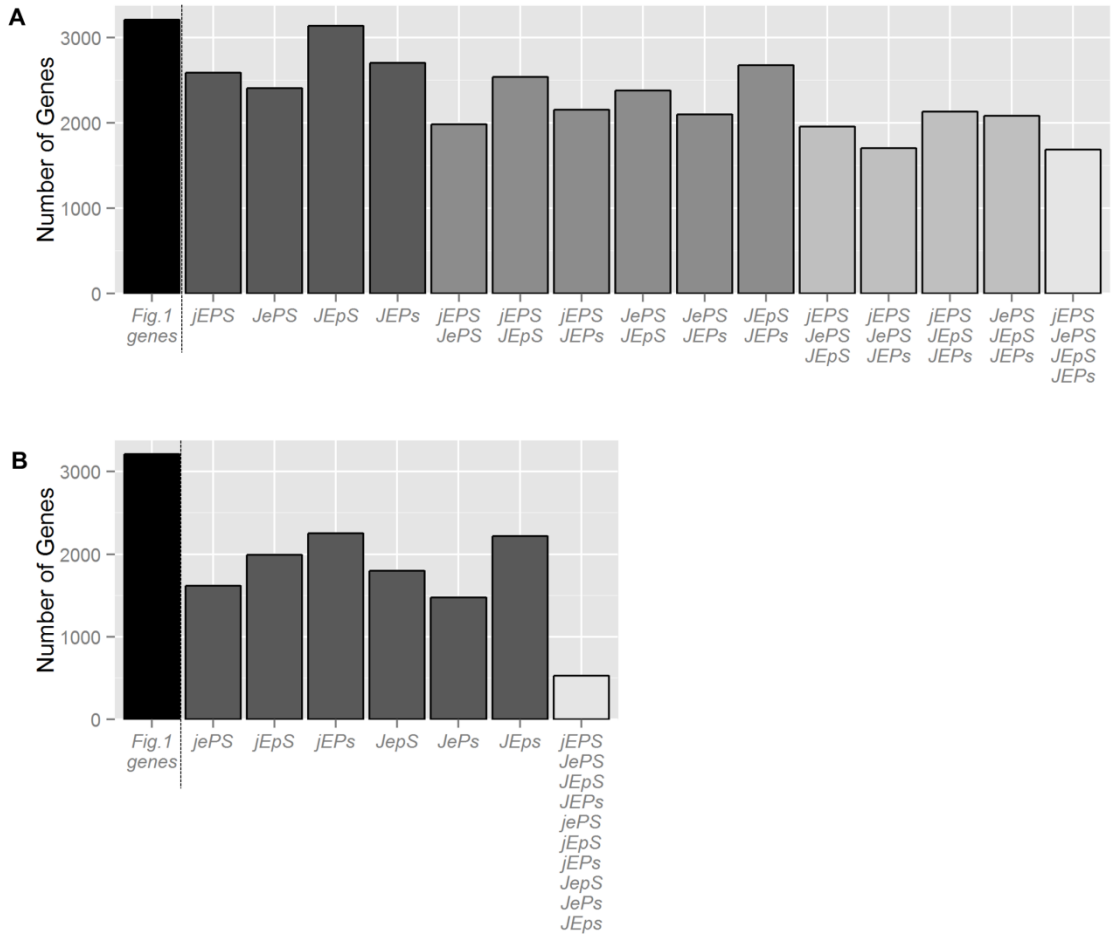


Figure 4-3. The responses of many genes to flg22 treatment are highly buffered by the network. Impacts of single sector perturbations are completely buffered in 1691 of the 3211 flg22-responsive, network-dependent genes. (A) Number of the 3211 genes with buffered transcript response changes in each single-sector perturbation, and in multiple single sector perturbations. The first bar represents the genes in Figure 4-1; the last bar represents genes buffered in all single sector perturbations (1691 genes). (B) Number of the 3211 genes with buffered transcript response changes in each double-sector perturbation, and (last bar) the number of genes buffered in all single and double-sector perturbations (527 genes). Combinatorial sector genotypes are labeled using a concise genotype notation that indicates the presence (uppercase) or absence (lowercase) of the sectors JA, ET, PAD4, and SA. *dde2-2*, *ein2-1*, *pad4-1*, and *sid2-2* were used to remove the JA, ET, PAD4, and SA sectors, respectively. For example, the *dde2-2 pad4-1* double mutant is denoted as *jEpS*.

4.3.4 The four sectors show contrasting patterns of network dependence

While the major immune hormones, SA and JA, are both involved in the immune response to flg22 stimulus, the ways in which these hormones are activated and modulated by the rest of the JA-ET-PAD4-SA network are profoundly different. Induction of free JA hormone levels was nearly entirely dependent on only the JA individual sector contribution (Figure 4-4A). Dependence on mainly the JA sector is also seen as a pronounced transcriptional pattern: a large set of 139 genes existed for which the main signaling allocation was JA alone (Figure 4-5A). In contrast, SA hormone levels could not be induced by flg22 independent of the other sectors; free SA was entirely dependent on the network for its induction (Figure 4-4D). Related to this, for no genes in the entire genome was the SA individual sector contribution the dominant signaling allocation (Figure 4-5D). The lack of such genes is no surprise: many immune genes are known to be SA-responsive, and thus we expect genes that respond to free SA hormone levels to also show, like free SA itself, complex network regulation. Compared to the transcriptome, free JA levels were among the least reliant on network interactions; the SA levels were among the most reliant on these interactions (arrows in Figure 4-2 indicate the histogram bins that contain JA and SA).

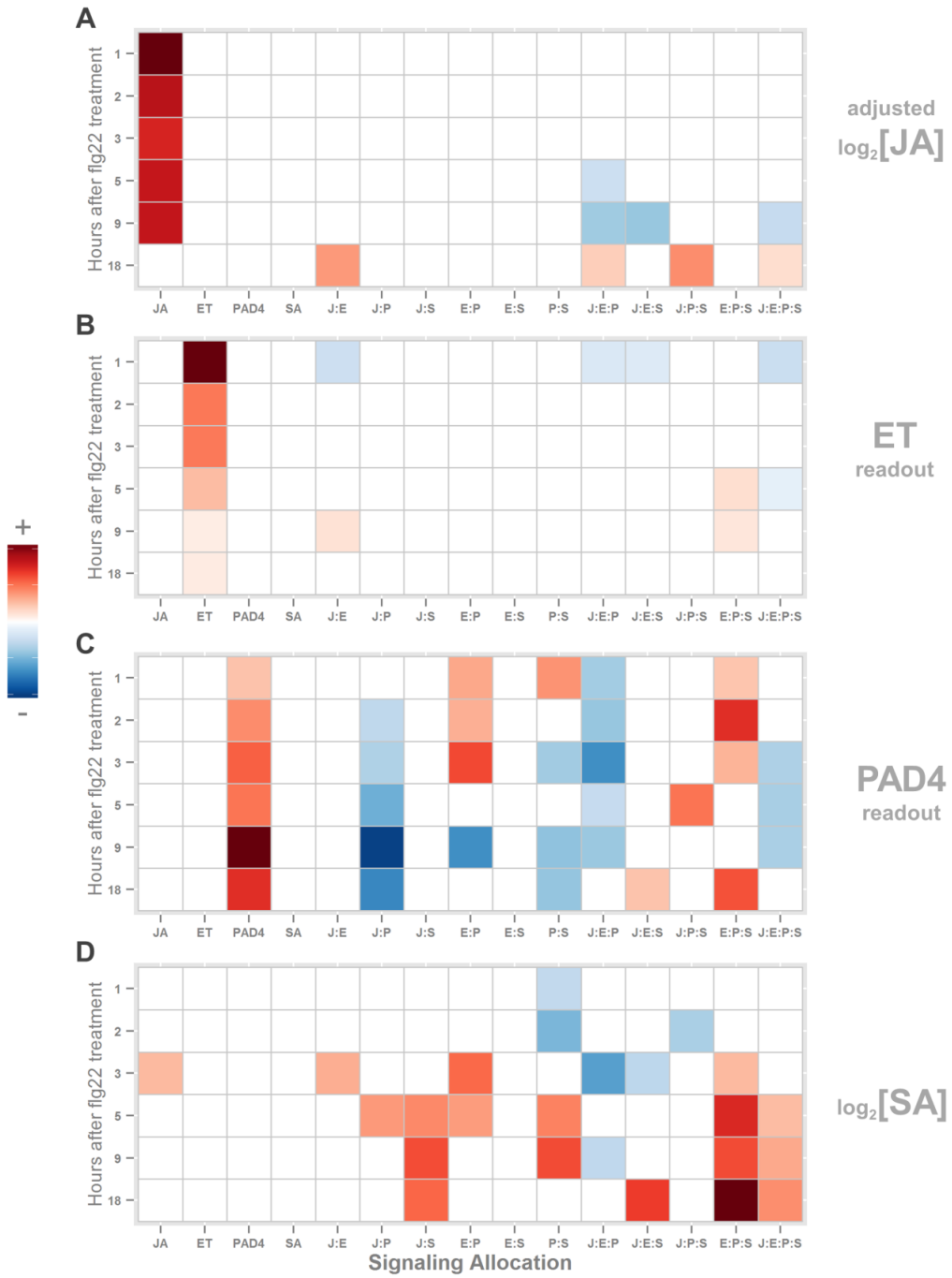


Figure 4-4. SA hormone induction completely depends on the other sectors. JA and ET hormone induction by flg22 is, on the other hand, mostly independent of the other sectors. Signaling allocations for responses of the following after flg22 treatment: (A) adjusted $\log_2$ free JA hormone concentrations, (B) transcript expression levels of the ET sector marker gene *ARGOS* (AT3G59900), (C) transcript expression levels of the PAD4 sector marker gene AT4504500, (D) $\log_2$ free SA hormone concentrations. The notations for the signaling allocations are the same as those in Figure 4-1. Only signaling allocations with q-values < 0.05 (Storey's FDR calculated across all 3211 responsive genes and 2 hormones) are shown. Hormones and genes were individually scaled for better visualization using the extreme value of each panel (sector).

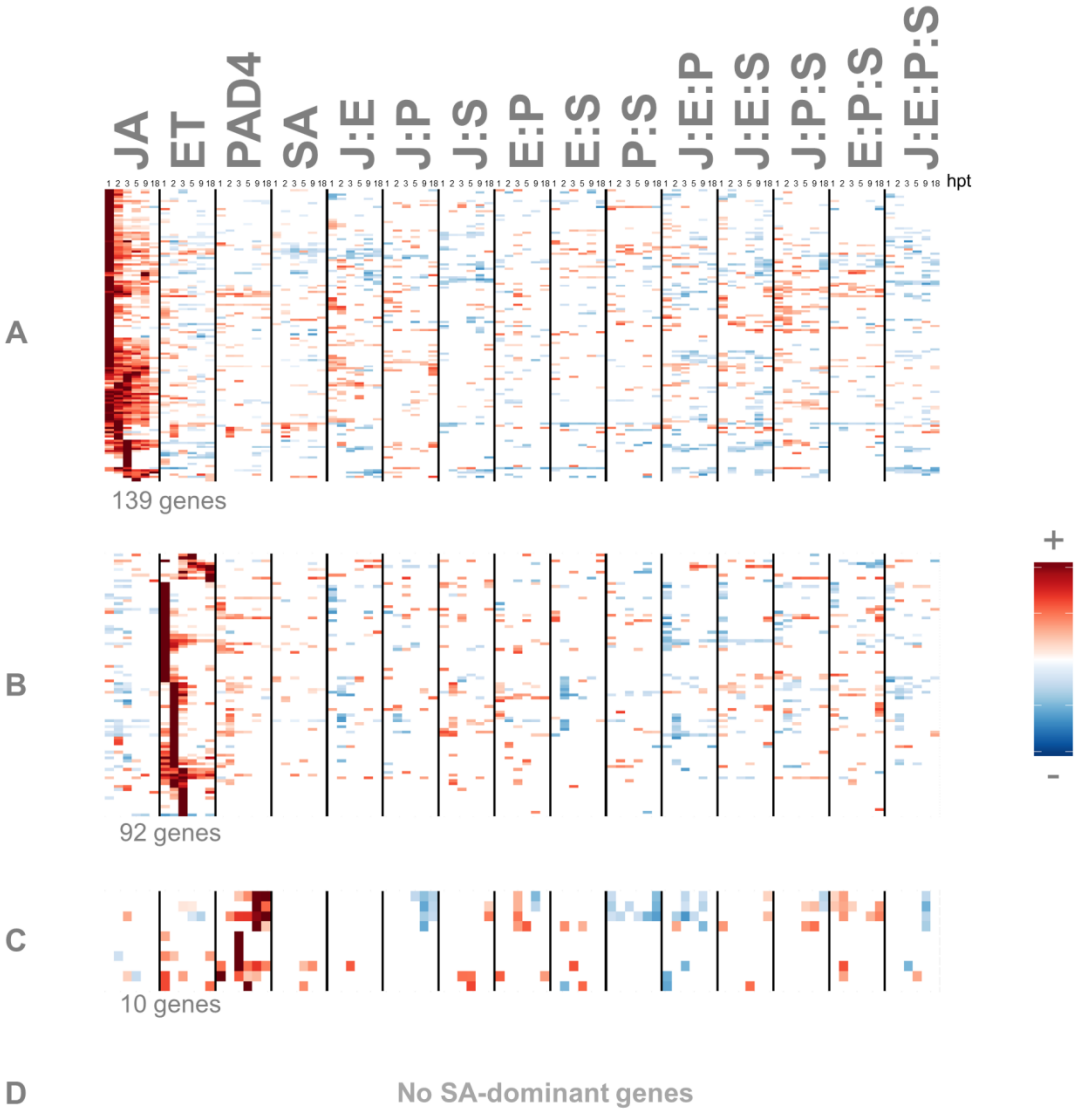


Figure 4-5. No genes are regulated by the SA sector alone. Genes whose transcript expression level responses are predominantly controlled by a single sector alone were selected for the (A) JA, (B) ET, (C) PAD4, and (D) SA sectors. For these genes, the maximum individual sector contribution is least 1.6 times greater than all other contributions to each signaling allocation. The sign and magnitude of the signaling allocation contributions are indicated by color and intensity; insignificant values are set to zero—visually they are white. Allocations were calculated for each gene or hormone from the transcript expression responses (abundance responses for hormones) through the time series 1, 2, 3, 5, 9, and18 hours post treatment (hpt); 6 minor columns representing the 6 time points post treatment follow each other consecutively within each major column. The notations of the contribution labels are the same as those in Figure 4-1. Only signaling allocation contributions with q-values < 0.05 (Storey's FDR) are shown. Rows were individually scaled for better visualization using the extreme value of each row.

Network regulations of the ET hormone response to flg22 resembled that of JA. Induction of ET hormone levels, shown by the transcript expression level of the ET sector activity marker *ARGOS* (AT3G59900), was nearly entirely dependent on the ET sector alone (Figure 4-4B). Similar results were seen with another ET sector activity marker ARL (S4 Fig). Transcript levels of both genes are likely reliable readouts of the plant response to the ethylene hormone since (I) they showed no response in any *ein2*-containing genotypes (*ein2-1* abolishes ethylene response phenotypes (Alonso et al. 1999)), (II) they are direct targets of the major transcriptional regulator of ethylene responses, *EIN3* (Chang et al. 2013), and (III) they both respond strongly and quickly to exogenously applied ACC (D. Winter et al. 2007; Markakis et al. 2012; Rai et al. 2015) (ACC is an ethylene biosynthetic precursor). As with JA, dependence on mainly the ET sector is a clear transcriptional pattern: 92 genes exist whose transcript response changes relative to the quadruple mutant response were almost entirely dependent on the ET individual contribution (Figure 4-5B).

Flg22 activation of the PAD4 sector showed a network dependence intermediate between those of SA and JA. PAD4 sector activation depended partly on the PAD4 individual sector contribution, but was also significantly affected by interactions among the sectors (Figure 4-4C). The PAD4 sector activity was monitored via the transcript expression level of a sector activity marker (AT4G04500) that passed strict criteria for abolished activity in *pad4*-containing genotypes. Other candidate PAD4 sector activity marker genes show similar network dependence (S5 Fig). We used these sector activity gene candidates to monitor the PAD4 sector since no method is known to directly monitor the activity of the PAD4 protein.

Activation of the four different summarized network sectors, JA, ET, PAD4, and SA, by flg22 treatment was largely consistent with our previous sector activity modeling results for MAMP-triggered PTI (Kim et al. 2014). Those results were: JA and ET are strongly activated by MAMP-responsive signaling, including flg22-responsive signaling; flg22 activates PAD4, but to a lesser extent than either JA or ET; flg22 activation of SA is entirely indirect: activation occurs through the other sectors of the network. Importantly, sector-interaction contributions to signaling allocations have two different sources: interactions within the 4-sector network, and those that occur downstream of

the network. Since free JA and the ET sector marker genes have simple signaling allocation signatures, a J:E interaction must involve signal integration downstream of the JA-ET-PAD4-SA network. For SA and PAD4, however, the sector activities are themselves regulated in a complex manner; interaction terms that include ‘S’ or ‘P’ could depend on interactions within the network, downstream of the network, or both. Note that genes that follow SA regulation (Figure 4-6 cluster D) are driven by complex network interactions, and yet the mechanistic interpretation is simple: these genes respond to free SA.

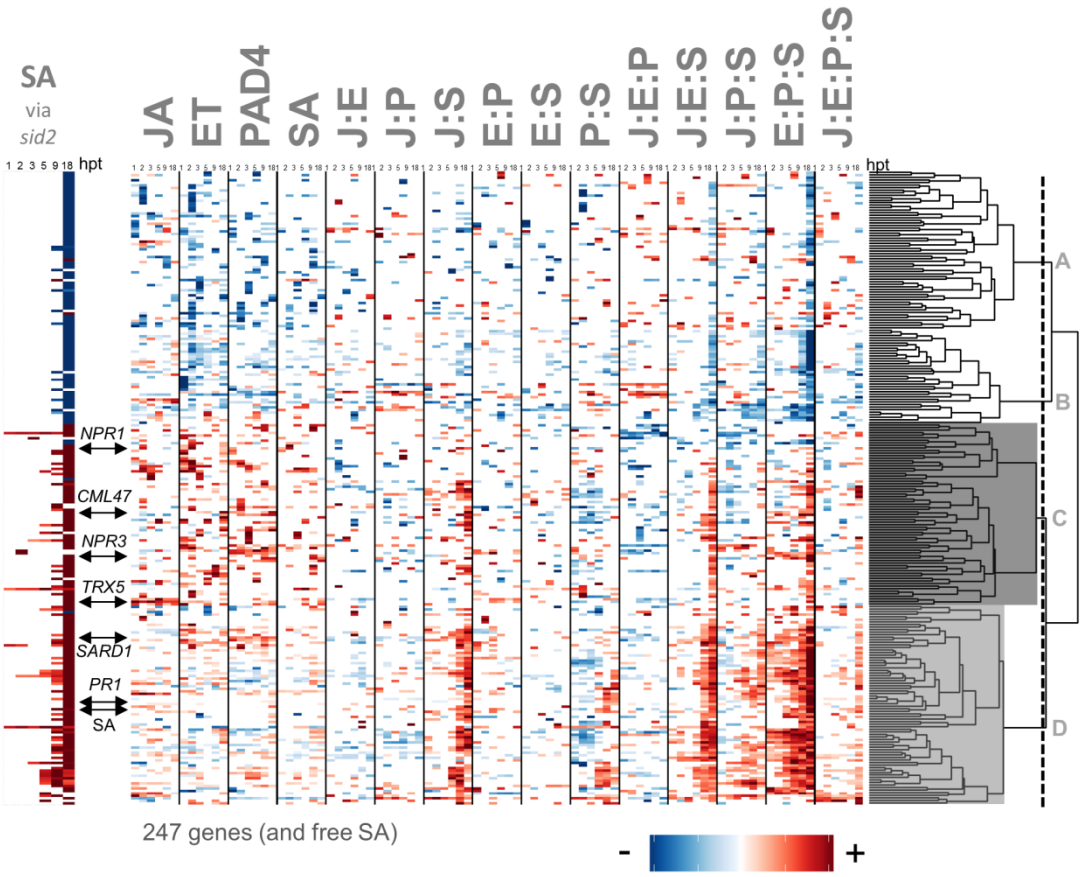


Figure 4-6. Two clusters of genes under distinct regulatory mechanisms are within the traditionally-defined SA-dependent genes. Signaling allocations are shown for the 247 genes that had a significant transcript expression level change in *sid2* relative to the transcript response in wild type. The left panel shows the strength of a traditionally-defined SA response: the opposite of the *sid2* mutant effect on the wild type response. For genes that are positively regulated by SA via this traditional definition, full signaling allocations revealed two distinct clusters: those genes whose regulation closely followed that of SA, and thus were regulated almost entirely by network interactions (cluster D, light gray) and those that also showed individual sector contributions (cluster C, dark gray). Canonical SA responsive genes *PR1* and *SARD1* are indicated, as well as *CML47*, a gene tightly co-expressed with *SARD1* across numerous pathogen treatments (Truman and Glazebrook 2012). Although the *SID2* gene

itself is in cluster C, note that the data for *SID2* transcript expression levels cannot be properly interpreted in the dataset since half of the combinatorial mutants have the *sid2* mutation. Additional genes well known for their involvement in SA-related plant processes, *NPR1*, *NPR3*, and *TRX5*, are also indicated. Only signaling allocation contributions with q-values < 0.05 (Storey's FDR) are shown. Rows were individually scaled for better visualization using the extreme value of each row.

4.3.5 Network reconstitution revealed unexpected patterns of transcriptional and hormone regulation

Antagonism between SA and JA is widely reported (Pieterse et al. 2012; Zheng et al. 2012; Van der Does et al. 2013). In contrast to these observations, we previously reported that induction of SA absolutely requires the JA sector after flg22 treatment when the PAD4 sector is absent (Kim et al. 2014). Here we report that, consistent with our previous discovery, both the SA hormone and the entire suite of genes whose regulation closely followed that of free SA (Figure 4-6, cluster D) have strongly positive J:S and J:E:S allocations at later time points (9 and 18 hpt). We note that the positive effect of JA on SA was entirely buffered by the network: SA induction was actually slightly higher in the single-sector mutant lacking the JA sector (*dde2-2*) than in the wild type (S6 Fig panel B). By traditional genetic analysis, this would have led us to infer that JA inhibits SA induction, in agreement with the literature. But network reconstitution allowed us to more accurately estimate the mechanisms underlying SA activation.

Although the PAD4 effect on SA inducibility is very similar to the JA effect on SA inducibility (S6 Fig panel B: the lines for *jePS* (green-blue) and *JepS* (red-blue) are very close), the PAD4 effect on the SA hormone abundance levels is higher than the JA effect on the SA hormone abundance levels because the PAD4 effect increases the basal SA level while the JA effect does not (S6 Fig panel A: compare *jePS* (green-blue) and *JepS* (red-blue)). The JA sector strongly and negatively regulates the PAD4 sector (Figure 4-4C: the J:P contribution to PAD4 sector activity is strongly negative). Thus, when the JA sector is removed from the wild type network, the PAD4 sector, which can increase the basal SA level, is relieved from its strong inhibition by the JA sector. As a consequence, the overall SA hormone level is increased by the PAD4 effect while the SA inducibility does not change much. An early study in *Nicotiana benthamiana* (Mur et

al. 2006) reported that plants treated exogenously with SA and low concentrations of JA showed strong positive synergy between SA and JA in the accumulation of *PR1a* transcript and protein (*PR1a* is the tobacco homolog of the Arabidopsis *PR1* gene, whose transcript expression levels are widely used as a proxy for SA hormone concentrations). This early observation can be explained by assuming that in *N. benthamiana*, or at the concentration of JA used in the study, the inhibition of the PAD4 sector activity by the JA sector activity is much weaker than for our conditions in *A. thaliana*.

PAD4 is required for SA hormone accumulation after pathogen challenge (L. Wang et al. 2011; Zhou et al. 1998). Yet the interaction between PAD4 and SA sectors, which we denote P:S, was a minor contribution to the signaling allocation for the SA hormone levels (Figure 4-4D) and to the transcriptome cluster that closely followed SA (Figure 4-6, light gray cluster). The visually dominant player was the high-order interaction E:P:S, meaning that a strong interaction between PAD4 and SA sectors required ET sector function as well. *FLS2* transcript expression levels were low in all *ein2*-containing mutants (S8 Fig), which is consistent with previous reports (Boutrot et al. 2010) and may explain this ET sector dependence in the E:P:S interaction. Yet *FLS2* transcript expression levels fully recovered in all *ein2*-containing mutants by 1hr after flg22 treatment, our earliest post-treatment time point (S8 Fig). In addition, activation of key kinases that mediate flg22 responses, MAP kinases *MPK3* and *6*, still clearly occurs in the quadruple mutant at a level comparable to the wild-type 10 min. after flg22 treatment (Tsuda et al. 2009). Thus, potentially reduced FLS2 protein concentrations in *ein2*-containing backgrounds likely did not have strong effects under our experimental conditions, probably due to treatment with a high concentration of flg22 (1 µM). The observed strong E:P:S interaction likely represents a genuine property of the network that mediates the signaling that ensues after FLS2 activation.

4.3.6 Network reconstitution revealed complex regulatory mechanisms that were hidden from null-mutant analysis of the individual genes in the 4-sector network

Induction of SA in response to pathogens is entirely absent in the *sid2-2* mutant (Wildermuth et al. 2001). As such, the *sid2-2* mutant, which we used to remove the SA sector, has been used in pathogen-responsive transcriptome profiling to define SA-dependent genes (L. Wang et al. 2008). We found that the true network regulation of SA-dependent genes that respond to flg22 stimulus could be classified into two distinct clusters (Figure 4-6 clusters D and C): (I) those whose regulation closely follows that of free SA and is entirely dependent on network interactions and (II) those that also show involvement of the individual sector allocations (JA, ET, PAD4 and SA). The GO annotations for these two positively regulated clusters overlap substantially: 80% of all individual annotations were shared between clusters (see S2 Table). Of the genes in these two clusters with mutant alleles that have been reported to show immune-related phenotypes in the literature (S1 Table), both clusters contain genes with SA-related mutant phenotypes. Network reconstitution, on the other hand, detects two distinct clusters.

Genes with significant transcript response changes in the ET single-sector mutant (*ein2-1*) showed two distinct signaling allocation patterns for both apparently positive and apparently negative regulation by the ET sector. Apparent regulation is based on the transcript response in *ein2* (Figure 4-7A, left hand column). Some apparent positive regulation by ET was true: many of the genes apparently regulated by ET were indeed regulated mainly by the ET sector alone (Figure 4-7A cluster 1). For a subset of genes apparently regulated positively by ET, however, the true regulation relied on a suite of interaction-based signaling allocation terms (Figure 4-7A cluster 2), similar to the set of interactions involved in SA hormone regulation, but also including the ET and PAD4 individual sector contributions. 20 of these genes overlapped with the SA cluster C genes (Figure 4-6), including WRKY53 and MLO2 (see S17 Text for the AGI locus identifiers of these 20 genes). Among genes with apparent negative regulation by ET, most were mainly negatively regulated by the ET sector (Figure 4-7A cluster 3), while for a distinct subset, the true regulation was a strong positive regulation by the JA sector, and only modest negative regulation by the ET sector (Figure 4-7A cluster 4).

For genes apparently regulated positively by the JA sector (Figure 4-7B), most of the true regulation was indeed positive regulation mainly by the JA sector alone (Figure 4-7B clusters 2 and 4), although more complex patterns were found, including a small set of genes that also showed reliance on several higher-order interaction terms that include ET as well as JA (J:E, J:E:P, J:E:S, and J:E:P:S; J:P:S was also implicated for these genes) (Figure 4-7B cluster 3). Quite surprisingly, genes which were apparently negatively regulated by JA showed a very different true regulation (Figure 4-7B cluster 1): strong positive regulation by ET and PAD4, as well as weaker regulation across many other signaling allocations. 27 of these 91 genes overlapped with the genes apparently under positive ET regulation (in Figure 4-7A), as would be expected for genes whose full regulation is strongly driven by the individual ET effect (see S18 Text for the AGI locus identifiers of these 27 overlapping genes).

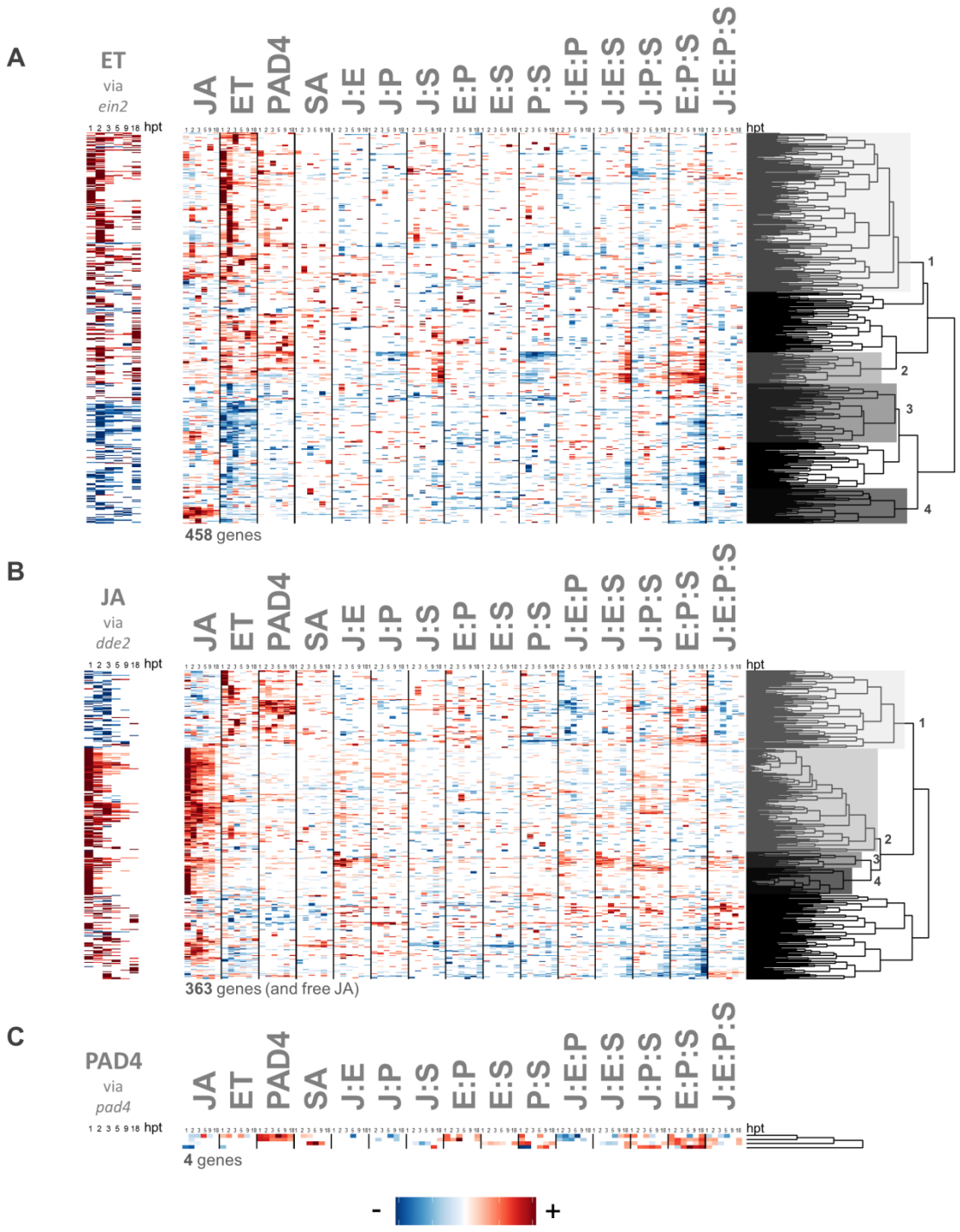


Figure 4-7. Regulation of genes strongly impacted by the JA and ET single sector perturbations. Network reconstitution uncovers the true regulation of genes whose flg22 transcript response changes are strongly impacted by the JA and ET single sector perturbations *dde2* and *ein2*. Signaling allocations are shown for the 458 genes (A) and 363 genes (B) that had a significant transcript expression level response in *ein2* (A) or *dde2* (B) relative to the transcript response in wild type. The left panels show the strength of a classically-defined sector response: the opposite of the single-sector mutant response. The full clustergrams show the true network

regulation of each of these sets of genes. Note the clusters that depend heavily on sector interactions (including A cluster 2, and B clusters 1 and 3), and those whose dominant allocation is not the individual contribution matching the relevant single-sector mutant (A cluster 4 and B cluster 1). Only 4 genes showed a significant transcript expression level response in *pad4* relative to wild type (C). Signaling allocation rows were scaled for better visualization using the extreme value of each row.

Only 4 genes showed significant transcript response changes in the *pad4-1* single-sector mutant relative to the wild type transcript response, as opposed to the hundreds of genes affected in the *dde2-2*, *ein2-1* and *sid2-2* single-sector mutants (Figure 4-7C). That is, loss of *PAD4* was almost entirely buffered in the flg22-responsive and network-dependent transcriptome. This is unexpected because *pad4-1* decreases flg22-PTI against *Pseudomonas syringae* pv. *tomato* (*Pto*) DC3000 by approximately 20% (Tsuda et al. 2009), and *pad4-1* has substantial pathogen growth phenotypes for many different pathogens (Glazebrook et al. 1997; Reuber et al. 1998; Zhou et al. 1998; Feys et al. 2001; Pegadaraju et al. 2005; Youssef et al. 2013). In our data, the *pad4* and wild type transcript expression levels are quite similar at each sampled time point; there is a trend for induced genes to be slightly less induced, and repressed genes to be slightly less repressed (S11 Fig), but these effects are small and few of them pass stringent cutoffs for differential expression. The *pad4-1* trend in gene expression is in agreement with the SA hormone concentrations relative to wildtype in *pad4-1* (S6 Fig A): mean SA hormone levels are somewhat lower in *pad4-1* (*JEpS*), but the effect size is much less than that of *sid2-2* (*JEPs*). It is possible that this strong buffering of the *pad4-1* single mutation effect may be specific to the response to flg22. We demonstrated that different MAMPs feed different patterns of input signals to the four-sector network to tune the network response (Kim et al. 2014). It is possible that buffering effects for loss of particular sectors may be modulated by differences in the signal flow in the network, which could be caused by different MAMPs and/or pathogen effectors. Consistent with this notion, the transcriptional response to *Pto* DC3000 *hrcC*, which is considered to present multiple MAMPs to plants, was affected similarly by the *pad4-1* mutation and the *sid2-2* mutation (Tsuda et al. 2008), suggesting that the induced SA level was very low in the *pad4-1* plant in that study.

As the transcriptional response to flg22 was barely affected by the *pad4-1* mutation, how could the mutation substantially decrease flg22-PTI measured with *Pto* DC3000? One

possible scenario is that flg22 treatment may result in non-transcriptional activation of a facilitator of PAD4's function. This potential facilitator activation ensuing flg22 treatment could enhance PAD4's function during infection. In this scenario, PAD4 would be required for the facilitator effect to be materialized, and consequently the *pad4-1* mutation would decrease flg22-PTI. Enhanced disease susceptibility 1 (EDS1) is a candidate for such a PAD4 facilitator since formation of an EDS1-PAD4 complex is essential for the function of this major regulatory component (Feys et al. 2001).

4.3.7 Genes in the ERF-branch of JA signaling did not display the expected strong synergy between JA and ET

One well-studied example of immune hormone cross-talk in Arabidopsis is a branch of JA signaling that depends on positive synergy between JA and ET. This ERF-branch requires both JA and ET for full activity (Pieterse et al. 2009), and is regulated by ethylene response factor (ERF) transcription factors (TFs). ERF-branch activity is commonly monitored by the marker genes *PDF1.2a*, *PR3* and *PR4* (Penninckx et al. 1998; Pré et al. 2008; Lorenzo et al. 2003; Verhage et al. 2011; Zhu et al. 2011). The ERF TFs are highly functionally redundant, making identification of ERF-branch genes by loss-of-function approaches difficult, but an overexpression line of the ERF TF gene *ORA59* was used to identify ERF-branch genes that are regulated by ERF TFs. By selecting within these genes for those that respond synergistically to exogenous JA and ethephon (an ethylene-releasing agent), a putative set of ERF-branch genes was assembled (Pré et al. 2008).

Since we observed concurrent spikes of JA and ET activity in our time course (see S7 Fig and S12 Fig for JA hormone levels and ET sector marker gene expression, respectively), our dataset is a natural place to test this set of candidate genes. Of the 37 candidate ERF-branch genes whose transcriptional response was stronger with a combined exogenous JA and ethephon treatment than with only exogenous JA treatment (Pré et al. 2008), 23 genes showed a significant transcript response change in our dataset. Nearly all of these 23 genes showed an unexpected lack of J:E allocations (Figure 4-8A). Notably, individual ET and to a lesser extent individual JA contributions

dominated the signaling allocations for these genes. This means that JA and ET have additive effects on the expression of these genes after flg22 treatment, but not synergistic effects. Unexpectedly, the canonical ERF-branch marker gene *PDF1.2a* was down-regulated (Figure 4-8B). *PR4* was down-regulated at early times, but then up-regulated at late times (S13 Fig); the network regulation during this induction, however, shows individual ET contributions, and no J:E contributions. The other canonical marker gene, *PR3*, was not significantly expressed in our data. The relatively low concentrations of JA observed here, as opposed to the very high JA concentrations used or observed in literature treatments during ERF-branch studies, may explain these unexpected results of the ERF-branch response. Additionally, the relatively short duration of the spikes in JA and ET activity (less than 2 hours), as opposed to the extended JA and ET induction in other immune responses, especially after inoculating plants with necrotrophic pathogens, may also have been a factor. We speculate that quantitative and/or kinetic differences in JA and ET hormone levels may produce different qualitative regimes for the J:E, J:E:P, J:E:S, and J:E:P:S terms.

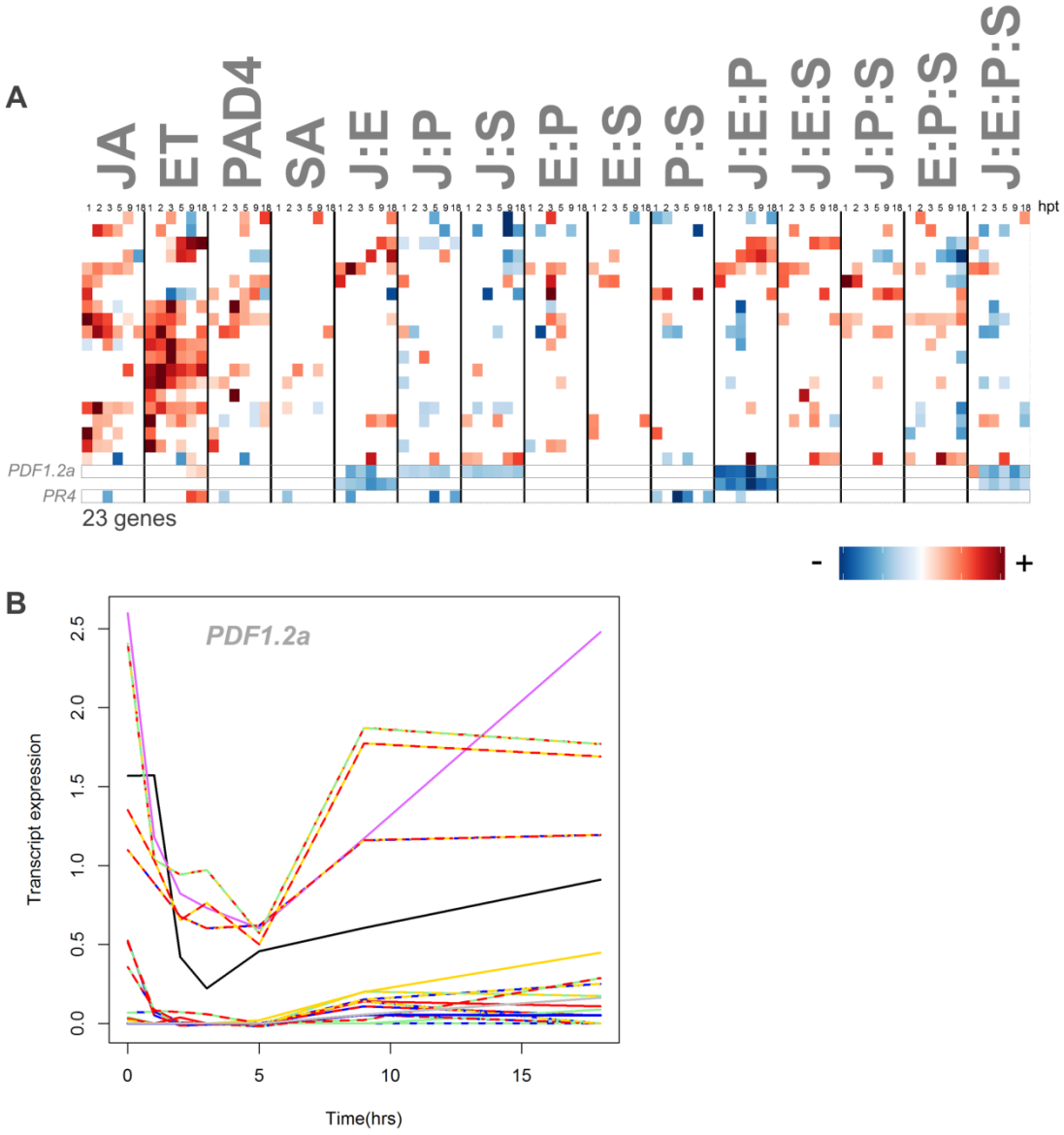


Figure 4-8. ERF-branch genes do not show the expected dominance of the J:E interaction. (A) Putative ERF-branch genes showed primarily JA and ET individual sector contributions. 37 differentially expressed genes from an ORA59 ERF overexpression line responded more strongly to a combined JA and ethephon treatment than to JA alone (genes selected using data in Table 1 from (Pré et al. 2008)). From this set of 37 genes, 22 genes were expressed in our dataset and showed at least one significant signaling allocation. These 22 genes are the intersect of the 37 genes and the flg22-responsive, network-dependent genes. Rows were scaled for better visualization using the extreme value of each row. (B) The classic ERF-branch marker gene, *PDF1.2a* is down-regulated after flg22 treatment. Mean transcript levels of *PDF1.2a* (AT5G44420) are shown across all genotypes and time points profiled. Colors indicate genotype by presence, not absence of a sector: red (JA), yellow (ET), green (PAD4), blue (SA). Other genotypes: black (wild type), gray (quadruple mutant), purple (*fls2*). Data were collected at 0, 1, 2, 3, 5, 9, and 18 hours after flg22 treatment.

As an alternative approach we selected flg22-inducible genes that show synergy between JA and ET (i.e., positive J:E interactions and non-negative or only very weakly negative JA and ET contributions) (Figure 4-9A). Of these genes, the two with the strongest J:E contributions, ANAC019 (AT1G52890, Figure 4-9B) and ANAC055 (AT3G15500, Figure 4-9C) were recently shown, together with ANAC072, to be unresponsive in an *aba2* mutant and directly up-regulated by the transcription factor MYC2, suggesting these genes belong in the MYC-branch of JA signaling which is known to show synergy between JA and the abiotic stress-response hormone abscisic acid (ABA). The MYC-branch has been shown to antagonize the ERF-branch (Pieterse et al. 2012). It will be interesting to investigate whether these genes with strong positive J:E signatures are regulated as ERF-branch response genes or whether they represent a different regulatory mechanism.

4.3.8 Promoter motif analysis of gene sets with similar signaling allocations

To further leverage our dataset to discover regulatory mechanisms underlying the plant immune response to flg22, we subjected the promoters (1 kb upstream and 5'UTR, if annotated) of sets of genes with similar signaling allocation signatures to MEME-based de-novo motif discovery (Bailey and Gribskov 1998). We did not find many significant motifs specific to particular regulatory patterns. However, we found broad involvement of GA-rich and GAA-rich motifs in flg22-responsive genes, which may be involved in chromatin structure and DNA integrity (see S5 Data and S2 Text for details).

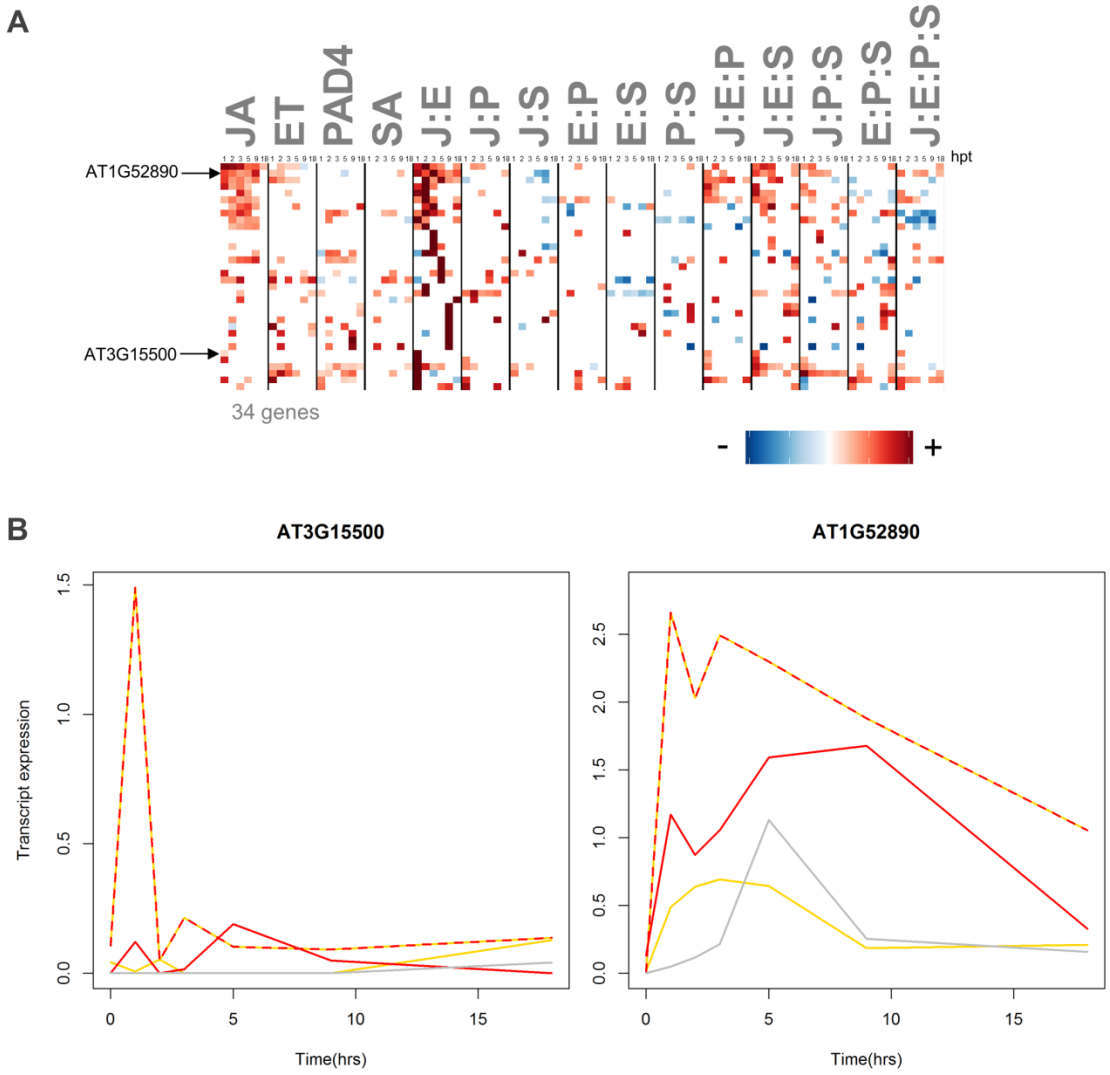


Figure 4-9. Several genes show dominance of the J:E interaction. (A) A clustergram of the signaling allocations for 71 genes that show a strong positive J:E interaction and non-negative or only very weakly negative JA and ET contributions in our dataset (JA, ET individual contribution > -0.25). Only signaling allocations with q-values < 0.05 (Storey's FDR) are shown. Rows were scaled for better visualization using the extreme value of each. (B) Mean transcript expression levels of the two of these genes with the strongest J:E contributions, *ANAC055* (AT3G15500) and *ANAC019* (AT1G52890), in the genotypes *JEps*, *Jeps*, *jEps*, and *jeps*, across time points profiled. Colors indicate genotype by presence, not absence of a sector: red (JA), yellow (ET); quadruple mutant expression is in gray. Data were collected at 0, 1, 2, 3, 5, 9, and 18 hours after flg22 treatment.

4.3.9 Conclusions

By definition, the components of a complex network are strongly connected. A systems biology approach allowed us to derive accurate mechanistic interpretations of component functions for a complex network where a classical genetics approach distorted interpretations of component function. In this paper we reduced a large complex network down to a manageable size—four major signaling sectors—to comprehensively probe, via network reconstitution, the sector interactions that regulate the transcriptome changes induced by a potent MAMP, flg22. Dynamic profiling of the transcriptome and two major immune hormones, SA and JA, across our set of comprehensive perturbation genotypes allowed us to describe and quantify complex mechanisms that drive the flg22 transcriptome response. Our analysis revealed that regulation of the vast majority of flg22-responsive and network-dependent genes rely on interactions between the JA, ET, PAD4, and SA signaling sectors. In many cases, such sector interactions provide a basis for network buffering, and in several cases the sector interactions driving transcript expression were unexpected. Our work shows that together high-order perturbation and dynamic transcriptional profiling can help in deciphering the regulation of a complex network, the plant immune signaling network. Network reconstitution is a powerful approach to reveal mechanisms that underlie the function of a complex biological network, especially one whose function is robust against pathogenic perturbations.

4.4 Materials and Methods

(See S1 Text for detailed method descriptions.)

4.4.1 Plant materials, growth conditions, leaf tissue treatment and collection

Combinatorial mutants (Tsuda et al. 2009) of the *Arabidopsis thaliana* mutants *dde2-2* (von Malek et al. 2002)*, ein2-1*(Alonso et al. 1999)*, pad4-1* (Jirage et al. 1999)*,* and *sid2-2* (Wildermuth et al. 2001), and the *fls2* mutant (SAIL_691C4) (Zipfel et al. 2004), all

in the *Col-0* background, were grown under a 12 hr photoperiod in a controlled environment at 22°C and 75% relative humidity. 3 leaves (leaf stages 7-9) of a single plant of each genotype from 31-32 days-old plants were syringe-inoculated with a 1 µM solution of flg22, and harvested at the indicated time after treatment. Different time point samples were collected from different plants. 0 hr samples were untouched. For each biological replicate, leaves of plants from four independent experiments (3 leaves per plant · 4 plants = 12 leaves in total) were pooled. Three biological replicates, from 12 independent experiments (4 plants per replicate · 3 replicates = 12), were profiled for each genotype:time combination, for a total of 17 genotypes·7 time points ·3 biological replicates = 357 tissue samples. (See S3 Text for detailed planting, treatment, and harvesting information.) Each tissue sample was divided for separate hormone measurement and Tag-Seq library preparation.

4.4.2 Hormone measurement

Tissue was flash frozen using N_2, powderized, and freeze-dried. Hormone extraction and concentration measurements were performed with an ultra-performance liquid chromatography-tandem mass spectrometer (UPLC-MS/MS) (ACQITY UPLC System/Quattro Premier XE; Waters) with an ODS column (ACQUITY UPLC BEH C18, 1.7 µm, 2.1 × 100 mm; Waters) as described previously (Kojima et al. 2009; Kojima and Sakakibara 2012). Samples were processed in two separate batches. Measurement values were corrected for the experimental factors $replicate$ and $batch$ using a mixed linear model for each hormone individually: $log_2[hormone] \sim genotype{:}\,time + (1|replicate) + (1|batch)$. Log_2 JA values near the detection limit were very noisy, and had a substantial offset from zero. This offset was subtracted and low values, including all values from *dde2*-containing genotypes, were set to zero, since no JA is produced without *DDE2* (Park et al. 2002). These modified data are referred to here as 'adjusted log_2 JA' values.

4.4.3 Tag-Seq library preparation and measurement of gene transcript expression levels

Stranded mRNA Tag-Seq libraries were prepared as previously described (Rallapalli et al. 2014). Internal barcodes were used to multiplex 16 different samples into a single lane of an Illumina flowcell, and lanes were sequenced using Illumina's Genome Analyzer IIx System. Only sequencing reads with exact barcode matches were used. Barcodes were trimmed, and reads were uploaded to Galaxy (Goecks, Nekrutenko, and Taylor 2010; Blankenberg, Von Kuster, et al. 2010; Giardine et al. 2005). Sequences that were mainly homopolymer sequence were removed using the Artifacts Filter in the fastx toolkit that is built in to Galaxy (Blankenberg, Gordon, et al. 2010). Remaining reads were mapped to TAIR10 transcripts using Bowtie (version bowtie-0.12.8) (Langmead et al. 2009). Mapping files for reads mapped to the forward strand were downloaded from Galaxy. Reads mapping to only a single gene were selected, and reads mapping to any transcript of a gene were summed together into a single gene-based count value. A scaling percentile normalization was applied to the raw counts of each sample: raw transcript counts at the 90th percentile of sample count values were scaled to match to the largest 90th percentile value across all samples, 300 counts. The data were transformed to produce approximately homoscedastic data values using the transformation $log_2(scaled\ counts + 79) - log_2(79)$. Normalized data were corrected for statistical artifacts using a mixed linear model for each gene individually: $normalized\ data \sim genotype{:}time + (1|machine/lane) + (1|barcode) + (1|replicate)$. The scaled, normalized and corrected data are called transcript expression levels here. The machine factor denotes that half of the samples were sequenced on a GAIIx machine at the University of Minnesota, and the other half on a different GAIIx machine at The Sainsbury Laboratory.

4.4.4 Gene selection

To find genes significantly regulated by the JA-ET-PAD4-SA network, p-values from each linear model were calculated for 90 difference of difference values per gene: $(e_{g,t} - e_{g,t=0}) - (e_{q,t} - e_{q,t=0})$ where e is normalized transcript expression, g is any non-quadruple combinatorial mutant (15 available), t is a time point after treatment (1, 2, 3, 5,

9 or 18 hpt) and q is the quadruple mutant genotype. These difference of differences values are called transcript response changes here. All 90 p-values (15 comparisons · 6 time points = 90 comparisons in total) from the 18,885 TAIR10 genes to which the mixed linear model for experimental factors could be fit were corrected together for multiple-hypothesis testing using Storey's FDR (Storey and Tibshirani 2003; Storey 2015). Q-values less than 0.05 produced by absolute transcript response changes greater than 1 were called significant. Genes with at least one significant transcript response change were considered significantly regulated by the network. For the cases (Figure 4-1 columns a and b; Figs 3, 6, 7) where genes were selected by transcript expression level responses relative to wild type, the wild type genotype was used as the baseline in the transcript response change calculations, instead of the quadruple mutant.

For how our infiltration stress control for the wild type response to flg22 treatment, *fls2*, behaves for both the 3211 genes in Figure 4-1 as well as the set of 6491 genes that show differential expression in either wild type or the quadruple mutant, see S3 Fig.

4.4.5 Signaling allocation models

For genes and hormones with significant transcript (or, for hormones, abudance) response changes, signaling allocation models were fit individually to each gene and hormone. These models were fit to transcript expression levels or hormone abundance levels relative to steady state, $e_{g,t} - e_{g,t=0}$, where g represents genotype, t time, and e transcript expression level (or, for hormones, abundance) (see S1 Text for details). Each starting model was regularized (that is, made sparse) using lasso (Efron et al. 2015; Tibshirani 2011). The lasso penalty parameter that minimized fit AICc (Hurvich and Tsai 1989; Burnham 2004) was selected. An ordinary linear model was then re-fit using the selected coefficient subset, to avoid the coefficient shrinkage inherent in lasso. P-values for all selected coefficients, for all genes and hormones relevant to a particular analysis (See S1 Text for details) were corrected using Storey's FDR. Coefficients for genes with insignificant q-values ($q \geq 0.05$) were set to zero, except in some Supporting Figs, where it is explicitly stated otherwise.

4.4.6 MEME-based motif discovery

The 3211 flg22-responsive, network-dependent genes and the hormones SA and JA were clustered based on their network signaling allocation models without remainder terms using complete-linkage agglomerative hierarchical clustering (see Figure 4-1). The resulting tree was cut at a height of 1.2, which produced visually-coherent sets of signaling allocations, resulting in 66 clusters. For the 57 clusters with more than 20 genes, the promoter regions of the genes in each cluster —1 kb upstream, and the 5'UTR, if annotated—were subjected to MEME-based motif discovery. Motif lengths of 8-15 were searched.

4.4.7 Clustering

The clustergrams and associated dendrograms in all figures were produced using cosine distance complete-linkage agglomerative hierarchical clustering.

4.4.8 Gene Ontology (GO) enrichment analysis

GO enrichment was performed using The Gene Ontology Consortium's web-based application (Ashburner et al. 2000; The Gene Ontology Consortium 2014). GO Ontology database release 2015-08-06; PANTHER Overrepresentation Test release 20150430.

4.4.9 Data submission to NCBI's Gene Expression Omnibus

Tag-Seq data has been submitted to NCBI's Gene Expression Omnibus (GEO); GEO data series GSE78735. Raw fastq files, tallied counts per gene for all samples, and pre-processed transcript expression level data are all available.

4.5 Acknowledgements

We thank the University of Minnesota Genomics Center (UMGC), and especially Aaron Becker, for help in sequencing the GAIIx libraries that were sequenced in Minnesota. We thank the Minnesota Supercomputing Institute (MSI) for data storage. We thank Jane

Glazebrook for useful suggestions, Mikiko Kojima for assistance with hormone measurement, and Barbara Kunkel for leading us to the (Mur et al. 2006) reference.

4.6 Supporting Information

4.6.1 S1 Text. Supporting Methods.

4.6.1.1 Plant leaf tissue treatment and collection

To handle the large number of samples in this study, plants harvested at all needed time points could not be infiltrated all at the same time. Thus inoculation was staggered so that different time points were inoculated at different times. The infiltration times were arranged so that the collection times were, as much as was experimentally feasible, approximately the same time of day. Specifically, leaves for the 1, 2, 3, and 5 hr time points were all harvested one right after another within about 2 hours. The 0, 9 and 18 hr samples were harvested at other times of day. See S3 Text for detailed infiltration schedules for all experiments.

4.6.1.2 Hormone measurement

4.6.1.2.1 Outlier detection

Data from the key hormone readouts used to track network response, JA and SA, were carefully evaluated for any outlier data points, since extreme outliers can have strong undesirable effects on linear model behavior. By plotting variance vs. mean for all sets of biological replicates (each point corresponds to a different genotype:time factor combination), a strong outlier was detected for JA hormone levels. The underlying point driving this outlier variance, *JEpS*, 18 hrs, replicate1, was masked with NA. A second outlier in the JA hormone levels was also found. There is no JA in *dde2*-containing genotypes (Park et al. 2002), but the JA measurements for these values bounced

around $\log_2(3)$ to $\log_2(7)$, a range which we considered to be the machine detection limit. One JA value, for *jepS*, 9 hrs, replicate 1, was near $\log_2(0)$. We considered this value an error, and masked it with NA. The mixed-effects linear model

$$log_2[hormone] \sim genotype{:}\, time + (1|replicate) + (1|batch) \qquad \textbf{Equation 4-1.}$$

was fit individually to 32 hormones (the other hormones did not have sufficient data in both batches to reliably fit this model). Replicate and batch effects were subtracted from $\log_2$[hormone] data to produce pre-processed hormone data (S4 Data). After determining that JA passed q-value based criteria for selection of flg22-dependent network-responsive genes and hormones, JA values were shifted so that the machine detection limit was about 0, as described in the main text, and data from *dde2*-containing genotypes were replaced with 0. Dramatic outliers were not seen in the SA measurements or in the Tag-Seq data.

4.6.1.3 Tag-Seq library preparation and measurement of gene transcript expression levels

Main text methods describe how raw gene transcript counts were tallied from sequencing data. See the NCBI Gene Expression Omnibus (GEO) Data Series GSE78735 for the sequencing data and sample metadata used in data processing. Tallied counts per gene in each sample, and pre-processed expression data are also available from this data series.

4.6.1.3.1 Scaling percentile-based normalization of gene counts from Tag-Seq data

We used a scaling normalization to make gene transcript counts from different samples comparable. 26,275 TAIR10 genes had at least one count in at least 1 sample. Only these genes were used in all further analyses. For each biological sample, gene counts were sorted, and the count value at the 90th percentile was extracted. The 90th percentile value for each sample was used to scale the counts in that sample to the sample with the largest 90th percentile count value, sample 150, with 300 counts at the 90th

percentile. For example, every gene count in Sample 1, which had the count value 42 at the 90th percentile, was multiplied by 300/42 (≈ 7.14). The distribution of all scaling factors used, for all 357 samples, is shown in S20 Fig. The top dozen genes with the highest expression levels were extremely variable and dominated the sequences in a given library, making library size normalization, a common RNA-Seq normalization procedure (Dillies et al. 2013), unreliable. Our percentile-based scaling normalization method is similar to library size normalization in that all the counts in an entire sample are multiplied by a single scaling factor. Our method differs only in how that scaling factor is determined.

4.6.1.3.2 Data transformation and correcting data for known experimental factors

We fit mixed-effects linear models to each gene (across all 357 values from each sample) to estimate the effects of genotype:time factors, and to correct for replicate effects and experimental artifacts (machine, lane and barcode) using the `lmer` function in the R package 'lme4' (Bates et al. 2015). For these models to be reliable, it was crucial that a major assumption of linear model fitting held true: the data needed to be approximately homoscedastic. Despite literature reports that a $\log_2$ transformation can be used for variance stabilization of RNA-Seq data from end-sequenced libraries (Rapaport et al. 2013), we did not find this to be the case. Variance in residuals for low-count values were well over an order of magnitude larger than those for high-count values. Residuals were extracted from the model

$$normalized\ data \sim genotype{:}\,time + (1|machine/lane) + (1|barcode) + (1|replicate).$$

Equation 4-2.

We thus selected a data transformation of the form

$$normalized\ data = \log_2(scaled\ counts + a) - \log_2(a)$$

Equation 4-3.

We searched integer values of a (to limit computational load; in principle a could be any constant) seeking values of a that minimized the variance range for the upper 75% of fit

mean values. Low fit mean values were excluded because we observed that values near zero had a variance near zero. a=79 was selected as the optimal a value (see S21 and S22 Figs).

The effects of experimental factors on scaled, normalized values were estimated using a separate linear model for each gene (Equation 4-2). The 357 transformed data values from all samples were used as input for each model. Zero-count data values were masked with NA for data correction purposes, since the artificial use of the value '0' can partially mask true patterns in the data. Genes with insufficient data to estimate replicate effects or machine effects were eliminated. For genes with insufficient information to fit all barcode effects, but with enough information to fit the other effects, the following model was fit:

$$normalized\ data \sim genotype{:}\, time + (1|machine/lane) + (1|replicate)$$

Equation 4-4.

For genes with insufficient information to fit all lane effects, the following model was fit:

$$normalized\ data \sim genotype{:}\, time + (1|machine) + (1|barcode) + (1|replicate)$$

Equation 4-5.

Genes which produced warnings or errors in fitting caused by insufficient data were removed from further analysis. 18,885 genes remained, and were used in the following analyses. Machine, lane, barcode, and replicate effects were subtracted from the scaled, normalized data. These corrected data are referred to as transcript expression levels in the main text.

4.6.1.3.3 Plotting mean expression for a given gene

NA values were replaced with zero ('0') when calculating mean values for all mean transcript expression plots (e.g. Figure 4-8B).

4.6.1.4 Gene selection

4.6.1.4.1 Calculation of transcript response changes

Many of the genes in whose expression we were particularly interested (e.g. PDF1.2a and PR1) had expression that included both very high values, as well as zero-count values, depending on the time point and genotype. Therefore masking zero with NA was an unsatisfactory way to deal with missing data when estimating transcript response changes. The true information about such data is that the value is less than or equal to 0, not that it is unknown. Thus, when estimating transcript response changes, we replaced NA with 0 (though admittedly this may underestimate the true transcript response change).

A transcript response change is a difference of differences; it calculates the transcript response in a network reconstitution genotype relative to the transcript response in the null network genotype, the quadruple mutant *dde2 ein2 pad4 sid2*:

$$r = \left(e_{g,t\neq 0} - e_{g,t=0}\right) - \left(e_{q,t\neq 0} - e_{q,t=0}\right) \qquad \textbf{Equation 4-6.}$$

Where r is the transcript response change, e the mean transcript expression level, g any combinatorial genotype other than q, the quadruple mutant; t is for time point. See S23 Fig for an example of how transcript response change values are calculated for the gene PR4 (AT3G04720).

4.6.1.4.2 Estimating q-values for transcript response changes

For each gene, 90 transcript response change values were estimated, for 15 genotypes and 6 responsive time points. P-values for each of these transcript response changes were calculated using the variance-covariance matrix of the gene's experimental effects linear model. When any one of the four transcript expression values could not be estimated in Equation 4-2 (because all three replicates were NA) that value was dropped

from the p-value calculation. Therefore, depending on the underlying data, a 4-sample, a 3-sample, a 2-sample or a 1-sample t-test was made. These 90 p-values for all 18,885 genes were corrected together with the 90 p-values of each hormone to which Equation 4-1 could be fit (32 hormones). Hormone level response changes were calculated in a manner analogous to how the transcript response changes were calculated: values under quantification ('U.Q.') were masked with NA, and since many hormones had a sizeable offset from zero, the lowest median replicate value was subtracted from the data of each hormone prior to p-value assessment. P-values for JA response changes were calculated from JA data with 2 masked outliers, but before replacing all data from *dde2*-containing genotypes with zero. Multiple-hypothesis testing correction to this entire set of p-values from Tag-Seq and hormone data was made using Storey's FDR (Storey and Tibshirani 2003; Storey 2015).Q-values less than 0.05 for transcript (or hormone) response changes greater than 1 or less than -1 were considered significant. 3214 genes had at least one significant q-value. The only active hormones with at least one significant q-value were JA and SA. All other genes and hormones were removed from further analysis. 3 of the 3214 genes did not have any significant non-remainder signaling allocation contributions. These 3 genes were also removed, leaving 3211 genes and 2 hormones for further analyses.

For the analyses in Figure 4-1 columns a and b, and Figs 3, 6, and 7, gene set selection (but not signaling allocation model calculation) was based on transcript response changes relative to the wild type for all 18,885 genes, instead of relative to the quadruple mutant genotype. For Figure 4-1 column a, and Figure 4-3A, p-values for comparisons of all single mutant transcript responses to wild type transcript responses were corrected together (4 genotypes · 6 time points = 24 p-values per gene). For Figure 4-1 column b and Figure 4-3B, p-values for all comparisons of single and double mutant transcript responses to wild type transcript responses were corrected together (10 genotypes · 6 time points = 60 p-values per gene). For Figs 6 and 7, p-values for all comparisons of the appropriate single mutant transcript responses relative to wild type transcript responses were corrected together (1 genotype · 6 time points = 6 p-values per gene). For Figs 6 and 7B, where the genotypes *sid2* and *dde2* were of interest, respectively, hormone abundance responses for SA and JA, respectively, were included with the p-values for the 18,887 genes for multiple-hypothesis testing correction. For Figure 4-1

columns a and b and Figure 4-3, we were interested in genes without transcript response changes relative to wild type. Thus genes were selected where all q-values were >0.05 and all transcript response changes relative to wild type were greater than -1 and less than 1. For Figures 4-6 and 4-7, seeking significant changes, we selected genes with at least one q-value <0.05 and an associated transcript response change relative to wild type greater than 1 or less than -1.

4.6.1.5 Signaling allocation models

4.6.1.5.1 Introducing noise into zero-count values for accurate estimation of signaling allocations

To estimate the signaling allocation contributions for each gene, the expression and hormone data were corrected for all random effects, and NA values were then replaced with zero (with the exception of the JA outlier for *JEpS*, 18 hrs, replicate1, which remained masked). Transcript or hormone responses were calculated by subtracting the corresponding t=0 value from the corresponding biological replicate and genotype. Replicate values of all 0 violate the homoscedastic assumption for fitting linear models, which include signaling allocation models. Thus noise was added to all transcript and hormone response zero-values before fitting each signaling allocation model. This noise was randomly drawn from a normal distribution with mean=0. For genes, the variance of this normal distribution was the average fitted residual variance for the top 75% of all mean transcript expression values across all 18,885 genes to which mixed-effects linear models were fit. For SA and JA, the residual variance was estimated separately for each hormone. Values were drawn at random from the appropriate one of these three distributions (one for genes, one for SA, one for JA) and these random values were used to replace the zero-count transcript response values. Random sampling can introduce artifacts, thus signaling allocation models were fit 100 different times for each gene, with a different random sample used each time from the normal distributions described above. The median value of each model output (signaling allocation contributions and associated p-values) was selected.

4.6.1.5.2 Fitting signaling allocation models

For each noise-added iteration, the linear model

$$\boldsymbol{e} = \mathrm{M}\boldsymbol{a} \qquad \textbf{Equation 4-7.}$$

was fit, where $\boldsymbol{e}$ is a vector of transcript expression (or hormone abundance) level response values of length 288 (16 network reconstitution genotypes across 6 induced time points, with 3 biological replicates for each genotype:time factor combination),

$$\boldsymbol{e} = \begin{bmatrix} \boldsymbol{e'}_{replicate\ 1} \\ \boldsymbol{e'}_{replicate\ 2} \\ \boldsymbol{e'}_{replicate\ 3} \end{bmatrix} \qquad \textbf{Equation 4-8.}$$

$$\boldsymbol{e'}_{replicate} = \begin{bmatrix} \boldsymbol{\eta_{t=1}} \\ \boldsymbol{\eta_{t=2}} \\ \boldsymbol{\eta_{t=3}} \\ \boldsymbol{\eta_{t=5}} \\ \boldsymbol{\eta_{t=9}} \\ \boldsymbol{\eta_{t=18}} \end{bmatrix} \qquad \textbf{Equation 4-9.}$$

$$\boldsymbol{\eta_t} = \begin{bmatrix} \boldsymbol{JEPS}_t \\ \boldsymbol{jEPS}_t \\ \boldsymbol{JePS}_t \\ \boldsymbol{JEpS}_t \\ \boldsymbol{JEPs}_t \\ \boldsymbol{jePS}_t \\ \boldsymbol{jEpS}_t \\ \boldsymbol{jEPs}_t \\ \boldsymbol{JepS}_t \\ \boldsymbol{JePs}_t \\ \boldsymbol{JEps}_t \\ \boldsymbol{jepS}_t \\ \boldsymbol{jePs}_t \\ \boldsymbol{jEps}_t \\ \boldsymbol{Jeps}_t \\ \boldsymbol{jeps}_t \end{bmatrix} \qquad \textbf{Equation 4-10.}$$

Where genotypes are indicated by the presence ($\boldsymbol{J}, \boldsymbol{E}, \boldsymbol{P}, \boldsymbol{S}$) or absence ($\boldsymbol{j}, \boldsymbol{e}, \boldsymbol{p}, \boldsymbol{s}$) of the four sectors (JA, ET, PAD4, SA).

$\boldsymbol{a}$ in Equation 4-6 is a vector of signaling allocation contribution coefficients of length 96, with 16 coefficients for each time point:

$$\boldsymbol{a} = \begin{bmatrix} \boldsymbol{\alpha_{t=1}} \\ \boldsymbol{\alpha_{t=2}} \\ \boldsymbol{\alpha_{t=3}} \\ \boldsymbol{\alpha_{t=5}} \\ \boldsymbol{\alpha_{t=9}} \\ \boldsymbol{\alpha_{t=18}} \end{bmatrix}$$ **Equation 4-11.**

$$\boldsymbol{\alpha_t} = \begin{bmatrix} \boldsymbol{JA}_t \\ \boldsymbol{ET}_t \\ \boldsymbol{PAD4}_t \\ \boldsymbol{SA}_t \\ \boldsymbol{J{:}E}_t \\ \boldsymbol{J{:}P}_t \\ \boldsymbol{J{:}S}_t \\ \boldsymbol{E{:}P}_t \\ \boldsymbol{E{:}S}_t \\ \boldsymbol{P{:}S}_t \\ \boldsymbol{J{:}E{:}P}_t \\ \boldsymbol{J{:}E{:}S}_t \\ \boldsymbol{J{:}P{:}S}_t \\ \boldsymbol{E{:}P{:}S}_t \\ \boldsymbol{J{:}E{:}P{:}S}_t \\ \boldsymbol{remainder}_t \end{bmatrix}$$ **Equation 4-12.**

M in equation 4-7 is a block matrix, composed of three M′matrices, one for each biological replicate:

$$M = \begin{bmatrix} M' \\ M' \\ M' \end{bmatrix}$$ **Equation 4-13.**

where M′is the block matrix:

$$M' = \begin{bmatrix} A & 0 & 0 & 0 & 0 & 0 \\ 0 & A & 0 & 0 & 0 & 0 \\ 0 & 0 & A & 0 & 0 & 0 \\ 0 & 0 & 0 & A & 0 & 0 \\ 0 & 0 & 0 & 0 & A & 0 \\ 0 & 0 & 0 & 0 & 0 & A \end{bmatrix}$$

Equation 4-14.

where 0 represents a block of zeros, and A the signaling allocation matrix defined in (Tsuda et al. 2009):

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/4 & 1/4 & 1/4 & 1/4 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1/3 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1/3 & 1/3 & 0 & 1/3 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Equation 4-15.

For the starting model, $\boldsymbol{e} = \mathrm{M}\boldsymbol{a}$, with only three times as many data points as coefficients fit, there were very few cases where any of these coefficients had significant p-values, even without correcting for multiple-hypothesis testing across all genes. Moreover, biologically, it is unlikely that all available signaling allocation contributions are used to regulate each gene. Therefore we sought sparse models that still have excellent predictive power.

Least-angle regression (lasso) regularization (Tibshirani 2011; Efron et al. 2015) provided a unique path for dropping model parameters. Lasso defined an order in which to drop parameters, which allowed us to avoid the combinatorial explosion of possible

paths for randomly dropping parameters. Lasso achieves regularization by using the L^1-norm times a shrinkage factor, λ, to penalize ordinary least-squares regression. Introducing this penalty has the effect of driving less significant parameters to zero, while shrinking all coefficients.

Using the `lars` function in the R package 'lars', coefficients were extracted for each step along the lasso shrinkage path. An ordinary least-squares model was then refit for each step, containing only the parameters with non-zero coefficients at that step. From each ordinary least-squares linear model fit along the lasso regularization path, Pearson correlation between the actual data and the model-predicted data was calculated. For a range of correlation cutoffs, from 0.8 to 0.99, in increments of 0.01, times the Pearson correlation of the full model, fitted coefficients (96 in total), p-values for each coefficient, and a model AICc (Hurvich and Tsai 1989; Burnham 2004) were extracted from the sparsest model with a Pearson correlation above the given cutoff. P-values for missing coefficients were set to 1.

For each gene or hormone, model fitting and regularization was repeated 100 different times, each with different randomly-sampled values from the zero-distribution replacing zero data values in each fit. Median values of each of the extracted model values (96 signaling allocation contributions, their associated p-values, and model AICc) across the 100 iterations were calculated for each correlation cutoff. Among these median-value models for each cutoff in the cutoff range, the model with the lowest median AICc was chosen. AICc is a form of Akaike's Information Criteria (AIC) that is corrected for finite sample sizes; using AICc is a standard statistical approach for balancing model complexity with model fit. After our lasso- and AICc-based regression, the models for many genes contained coefficients with highly-significant p-values.

Lasso- and AICc-regularized model coefficients were calculated separately for each gene; for most analyses the 90 p-values for non-remainder signaling allocations extracted from these models were corrected for multiple-hypothesis testing together for all 3211 genes and 2 hormones. For the analyses in Figs 6 and 7, p-values were corrected for all signaling allocation contributions (including the remainder) for the set of

genes and hormones in each figure panel (Figs 6, 7A, 7B, 7C), since these gene sets were not entirely subsets of the 3211 genes and 2 hormones. The `qvalue` function in the R package 'qvalue' (Storey 2015) was used to correct all p-values simultaneously using Storey's FDR method (Storey and Tibshirani 2003). P-values set to 1 were removed from the set of p-values before correction; a histogram of all non-1 p-values was visually inspected to ensure that the assumed null distribution of Storey's FDR method was met for our calculated p-values. It was met for every gene set tested in this paper. Coefficients of parameters with q-values less than 0.05 were set to zero. Except where noted otherwise, only coefficients with significant q-values were used in graphical model representations (clustergrams) and any subsequent calculations that used model parameters.

As stated above, signaling allocation models could not be fit directly to the expression data because of the prevalence of zero-count values. These values skew the variance distribution for any gene, especially when all three replicate values for a given genotype:time combination are zero; badly behaved variance distributions that are far from homoscedastic would have resulted in unreliable fits. Hence we used the signaling allocation contribution coefficient and q-value selection procedure discussed above. It is possible, however, to assess correlation between data and predicted data that were predicted from regularized models. The distribution of these correlations for the 3211 genes and 2 hormones in Figure 4-1 are shown in S24 and S25 Figs. The median Pearson correlation values before and after q-value based regularization were 0.89 and 0.84, respectively.

4.6.1.6 Selection of sector activity marker genes for the ET and PAD4 sectors

The first selection criterion for sector activity marker genes was that the transcript expression levels had to be unresponsive in the quadruple mutant. The linear models (Equations 2, 4, and 5) used to estimate the effects of experimental factors were used to calculate values and p-values for all 21 differences in quadruple mutant expression levels between any two of the seven time points. P-values for these 21 differences across all 18,885 genes were corrected together using Storey's FDR. 145 genes were

found where all differences were greater than -1 and less than 1, and where all q-values were less than 0.05. Of these, genes that also had at least one significant transcript response change relative to the quadruple mutant (i.e. genes that were also among the 3211 genes selected for Figure 4-1) fulfilled the first criterion for sector activity marker genes.

For PAD4 sector activity markers, genes were selected for which the total absolute value of all significant allocation contributions excluding the PAD4-containing contributions (e.g. E:S does not contain P) summed to less than 1. Further, genes were required to have at least one significant PAD4 individual sector contribution across the time course. Genes with P:S contributions but without any PAD4 individual sector contributions were eliminated. This left 6 candidate genes: AT4G04500, AT5G24200, AT4G21840, AT2G18660, AT1G21525 and AT1G05880. AT1G05880 was eliminated since the PAD4 individual contributions were significant but very weak compared with the other significant PAD4-containing contributions.

For ET sector activity markers, genes were selected for which the total absolute value of all significant allocation contributions that did not include the ET allocation (e.g. P:S does not contain E) summed to less than 1. Genes were also required to have at least one significant ET individual sector contribution somewhere in the time course. Further, genes were required to be targets of the transcription factor EIN3 (Chang et al. 2013). Moreover, genes were required to show strong responses to ACC treatment in the Arabidopsis eFP browser (D. Winter et al. 2007), including at the earliest time point available in the browser, 30 min. after treatment. Finally, genes were required to be specifically referenced in the literature as ethylene-responsive. This left only two candidates, ARGOS (AT3G59900) and ARL (AT2G44080). Since the genes have similar annotation, the ACC transcriptional response readout EBF2 (AT5G25350) used in the literature to report ethylene responses (Guo and Ecker 2003) was also used as supporting evidence to argue that a spike in ethylene hormone levels likely occurs *in planta* after flg22 treatment.

4.6.1.7 MEME-based motif discovery

Comparisons of the de-novo motifs we found to the motifs in CIS-BP (Weirauch et al. 2014) and AtcisDB in AGRIS (Davuluri et al. 2003; Yilmaz et al. 2011) were made using the Kullback-Leibler distance. Per literature recommendation (Aerts et al. 2003), distances less than 0.4 were considered matches. See Supporting Data for all discovered motifs and database matches. No motifs matched those in the AtcisDB database.

4.6.2 S2 Text. Motif Analysis Summary.

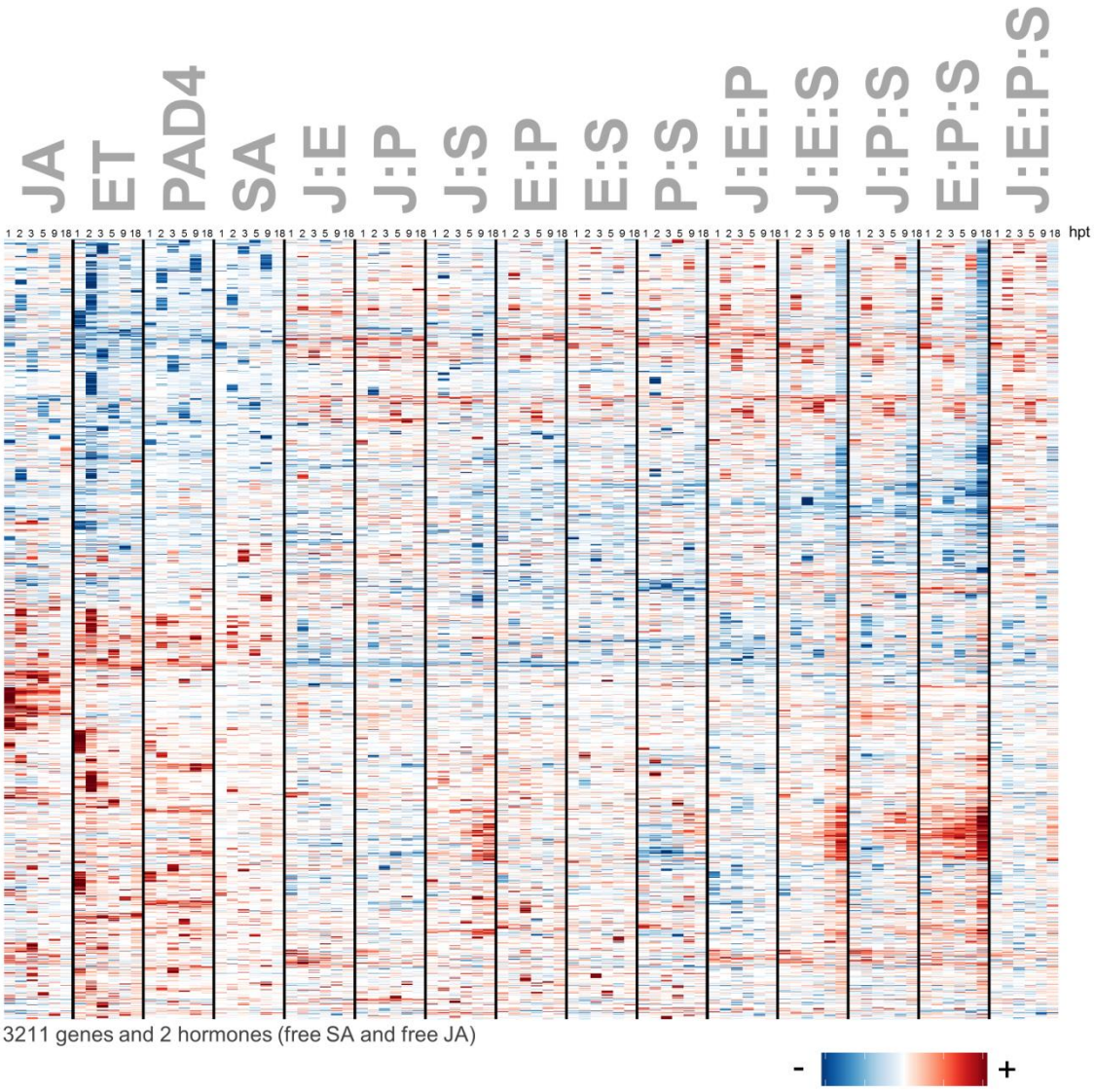
4.6.2.1 Promoters of flg22-responsive, network-dependent genes are highly enriched for GA-rich and AAG-rich motifs

To further leverage our dataset to discover regulatory mechanisms underlying the plant immune response to flg22, we subjected the promoters (1 kb upstream and 5'UTR, if annotated) of sets of genes with similar signaling allocation signatures (S6 Data) to MEME-based de-novo motif discovery (Bailey and Gribskov 1998). Nearly every tested cluster showed enrichment for GA-rich motifs and/or the reverse complement CT-rich motifs (example in S19 Fig A, full motif sets in S5 Data). Plants appear to have two distinct sets of genes: those whose transcription is initiated at TATA-boxes, and those enriched for GA-rich sequences (Molina and Grotewold 2005). GA-rich motifs have been found in Arabidopsis promoters of targets of the developmental transition regulator *LFY* (C. M. Winter et al. 2011) and targets of the circadian master regulator *TOC1* (Gendron et al. 2012). Further, GA-rich motifs have been linked to chromatin modification in Arabidopsis: the Basic Pentacysteine1 protein binds GA-rich motifs (Hecker et al. 2015), putative GAGA binding sites are enriched in Polycomb Repressive Complex 2 binding sites (Deng et al. 2013), and GAGAGA motifs are enriched in target loci of the Heterochromatin Protein1 Arabidopsis homolog *LHP1*, and in target loci of H3K27me3 (Hecker et al. 2015). Chromatin modification has been implicated in Arabidopsis innate immunity (Ding and Wang 2015), but chromatin involvement in flg22 responses or

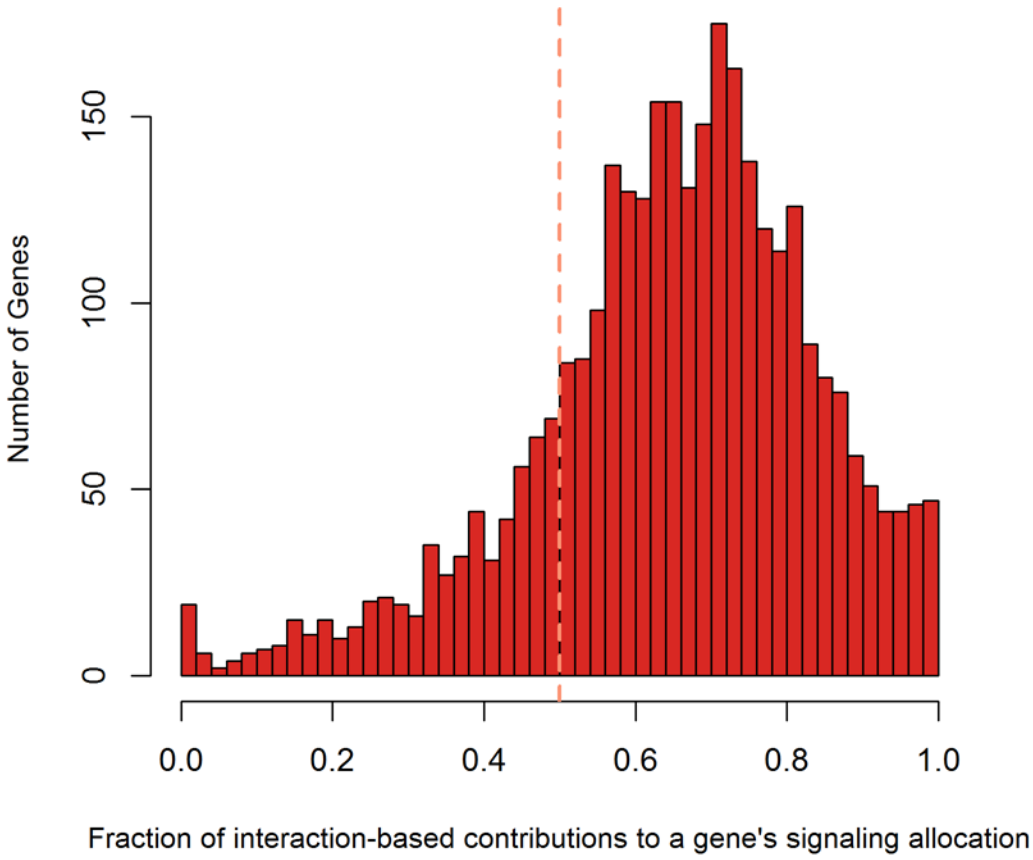
broadly in PTI has not yet been shown, to the best of our knowledge. It will be interesting to see what role these GA-rich motifs play in flg22 responses and whether chromatin modification may play a role in PTI.

The other major motif patterns discovered in the promoters of nearly every tested cluster were AAG-repeat motifs (and/or the reverse complement CTT-repeat motifs) (example in S19 Fig B, full motifs sets in S5 Data) and A-rich motifs (and the reverse complement T-rich motifs) (see S5 Data). Pajerowska-Mukhtar et al. (Pajerowska-Mukhtar et al. 2012) found that the motif GAAGAAGAA was enriched in flg22-responsive genes, and that the transcription factor TBF1 binds to DNA fragments enriched for this motif. Additionally, recent work has shown that CTT-repeat motifs are the most highly enriched motifs in recombination hotspots in the Arabidopsis genome, followed by A-rich motifs (Choi et al. 2013). The fragility of $(AAG)_n$ sequences may be due to unusual 3-way base pairing possible with $(AAG)_n$ sequences, the Hy3-type intramolecular triplex (Sinden et al. 2002) that leaves one DNA strand unpaired. It has recently been shown that SA directly damages DNA and that this damage potentiates Arabidopsis immunity (Yan et al. 2013) and that BRCA2 and RAD51 are directly involved in plant immunity (S. Wang et al. 2010). It will be interesting to see what role these AAG-repeat and A-rich motifs play in the plant immune response.

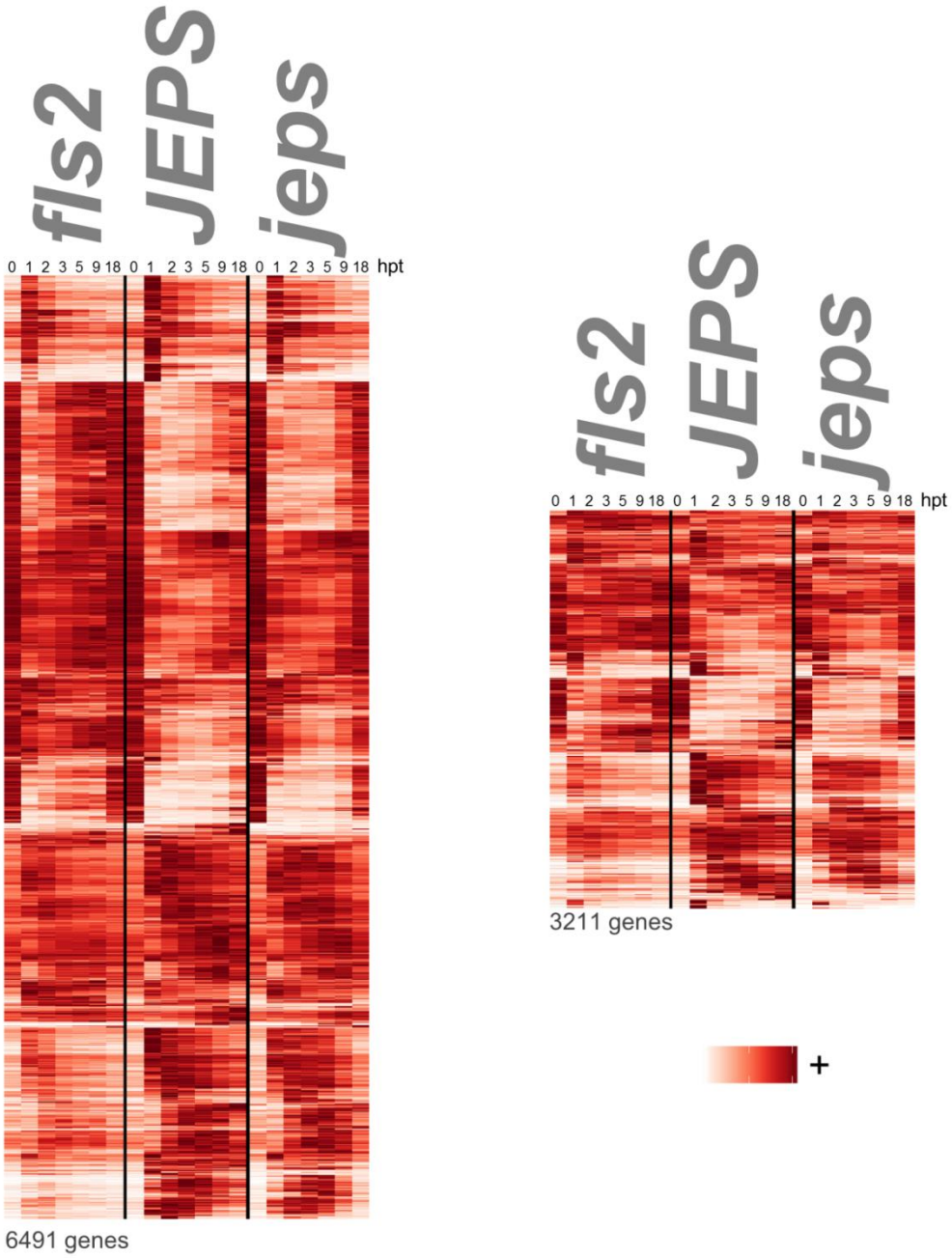
4.6.3 Supporting Figures and Tables



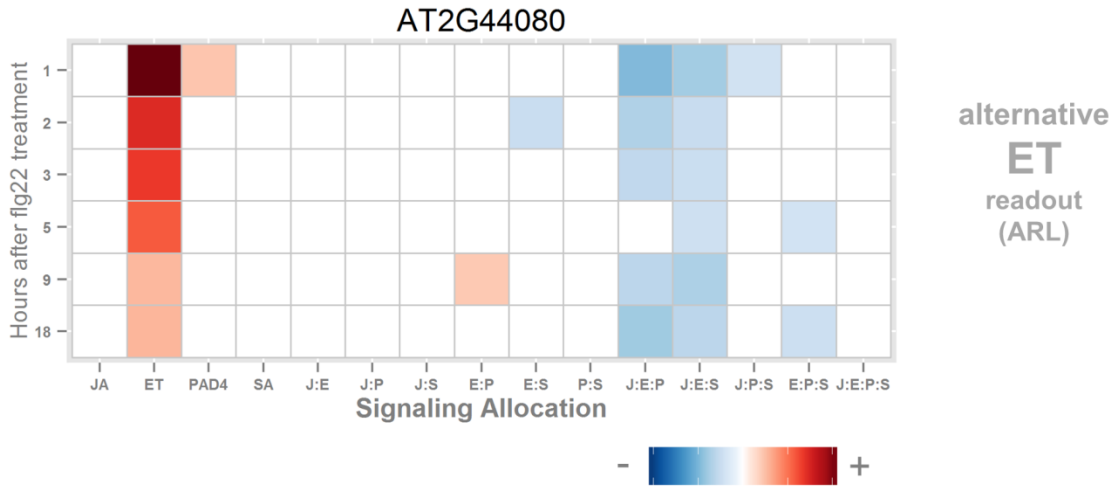
S1 Fig. Signaling allocation models for Figure 4-1 genes and hormones without q-value-based regularization. This figure is similar to Figure 4-1, but shows all coefficients left from the first lasso- and AICc-based step of our 2-step regularization of signaling allocations. The row ordering of genes and hormones here is the same as in Figure 4-1.



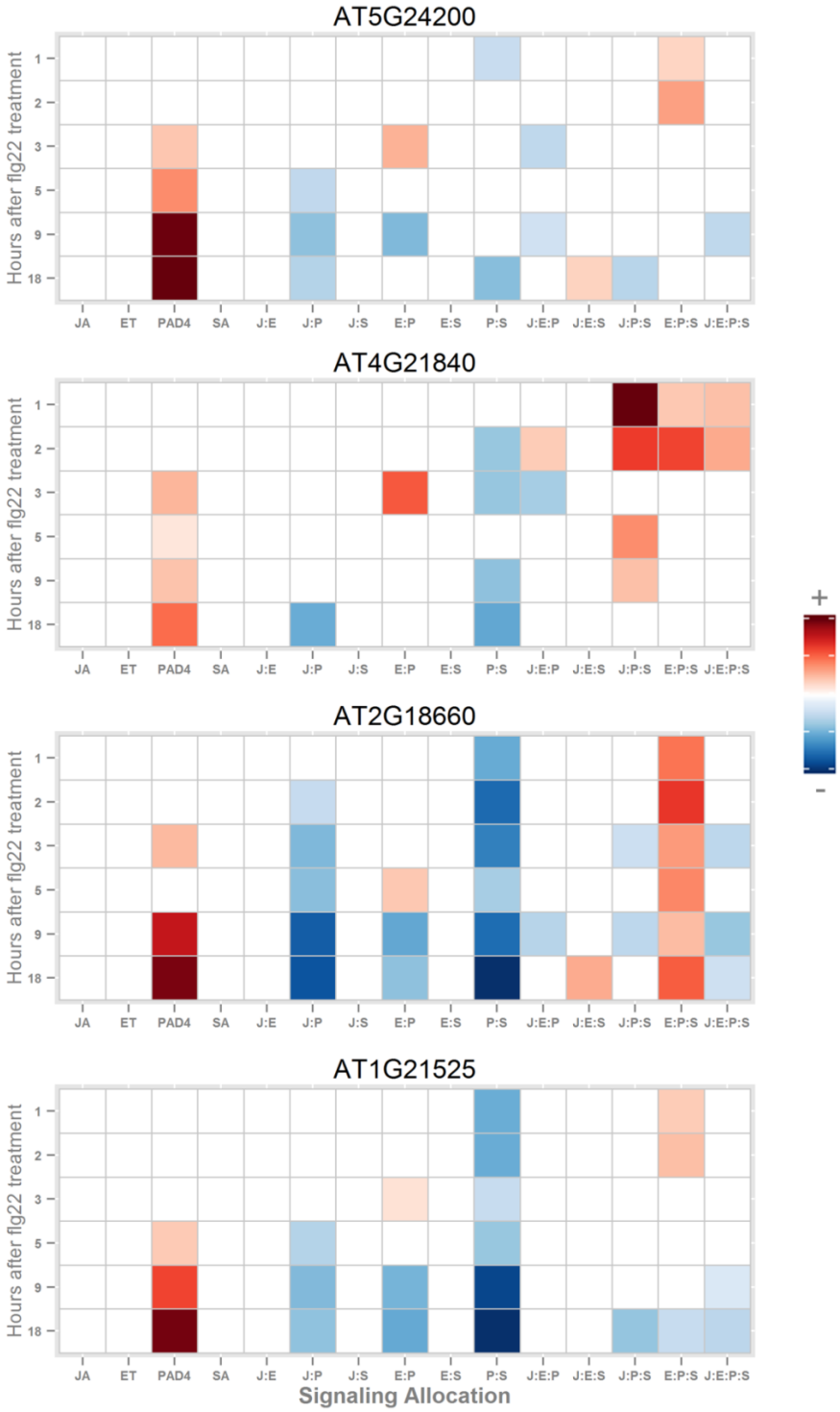
S2 Fig. Fraction of interaction-based contributions to the signaling allocations of genes and hormones in S1 Fig. This figure is similar to Figure 4-2, except that no q-value based regularization was applied to the signaling allocations. For each gene or hormone, the fraction was calculated via the Euclidean length of a vector representing only the sector interaction terms divided by the Euclidean length of the vector representing the full signaling allocation.



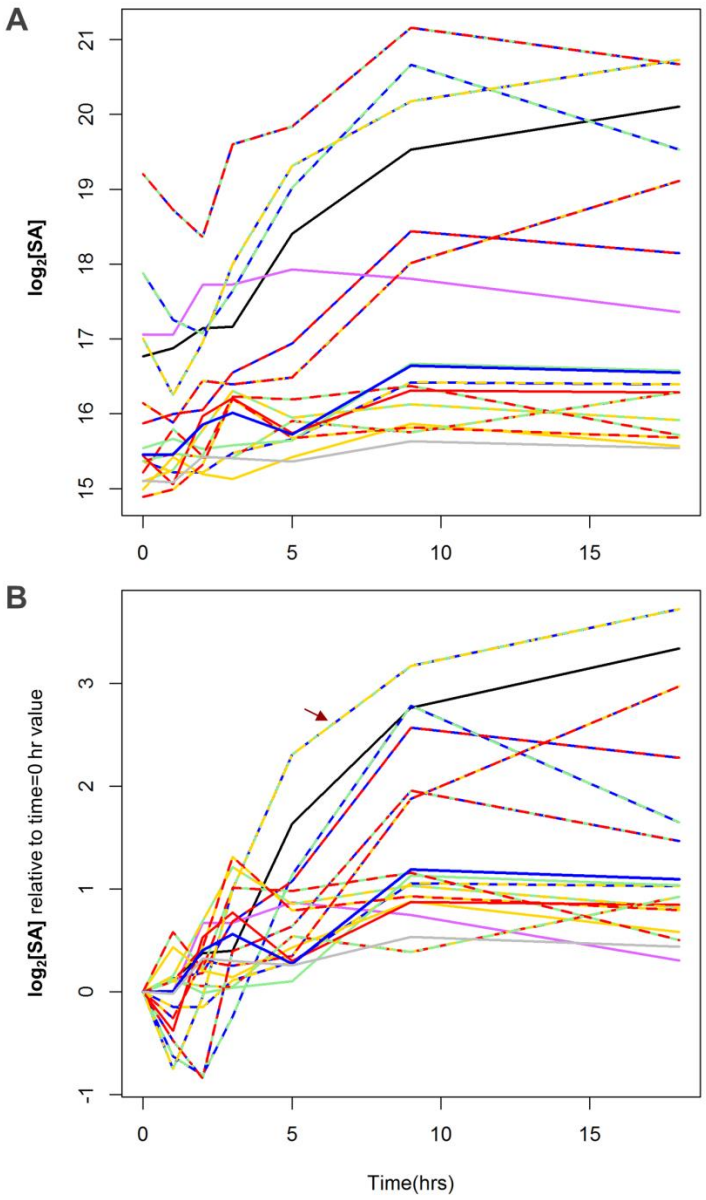
S3 Fig. Transcriptional behavior of flg22-responsive genes in *fls2*, the wild type, and the quadruple mutant. The mean transcript expression levels in the *fls2* genotype, compared to the wild type (*JEPS*) and the quadruple mutant (*jeps*), is shown for (A) all genes with differential expression relative to the 0 hr expression value in either *JEPS* or *jeps*. Differentially expressed genes are those whose induction or repression is greater than 1 or less than -1, respectively, and with associated q-values < 0.05. (B) Mean transcript expression levels of the 3211 genes in Figure 4-1. Note that the number of genes in Figure 4-1 represents roughly half of the flg22-responsive genes. Mean transcript expression is shown across the sampled time course 0, 1, 2, 3, 5, 9, 18 hours post treatment (hpt) with flg22; rows were individually scaled for better visualization using the extreme value of each row.



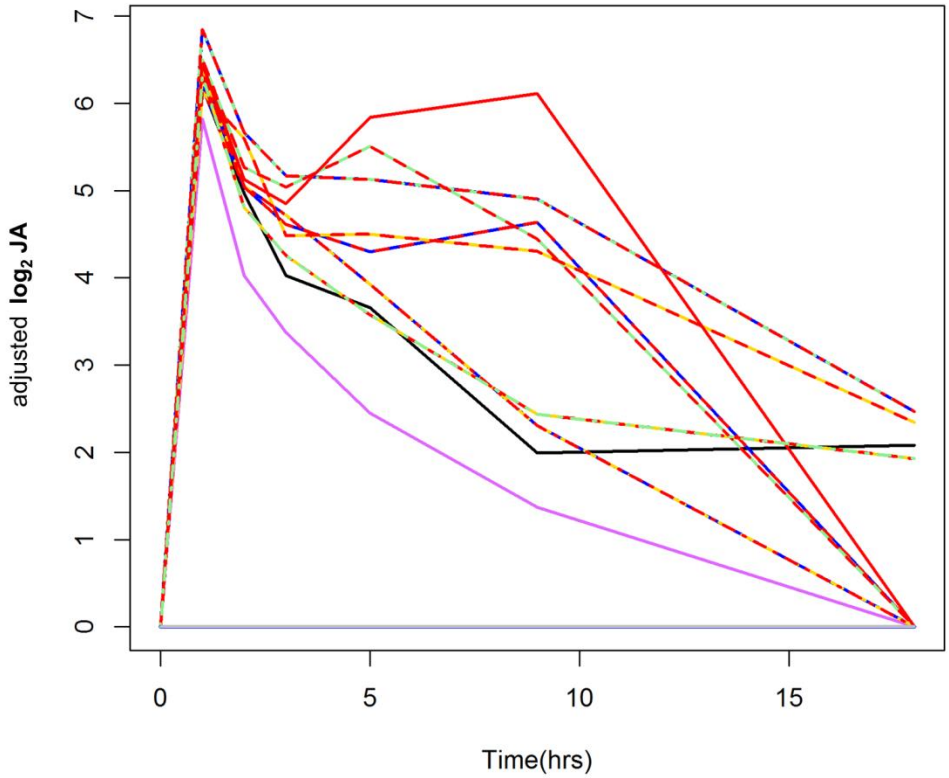
S4 Fig. Signaling allocation for an alternative ET sector activity marker, *ARL* (AT2G44080). Only signaling allocation contributions with q-values < 0.05 (Storey's FDR) are shown. The allocation was scaled for visualization using the extreme value (ET, 1 hr).



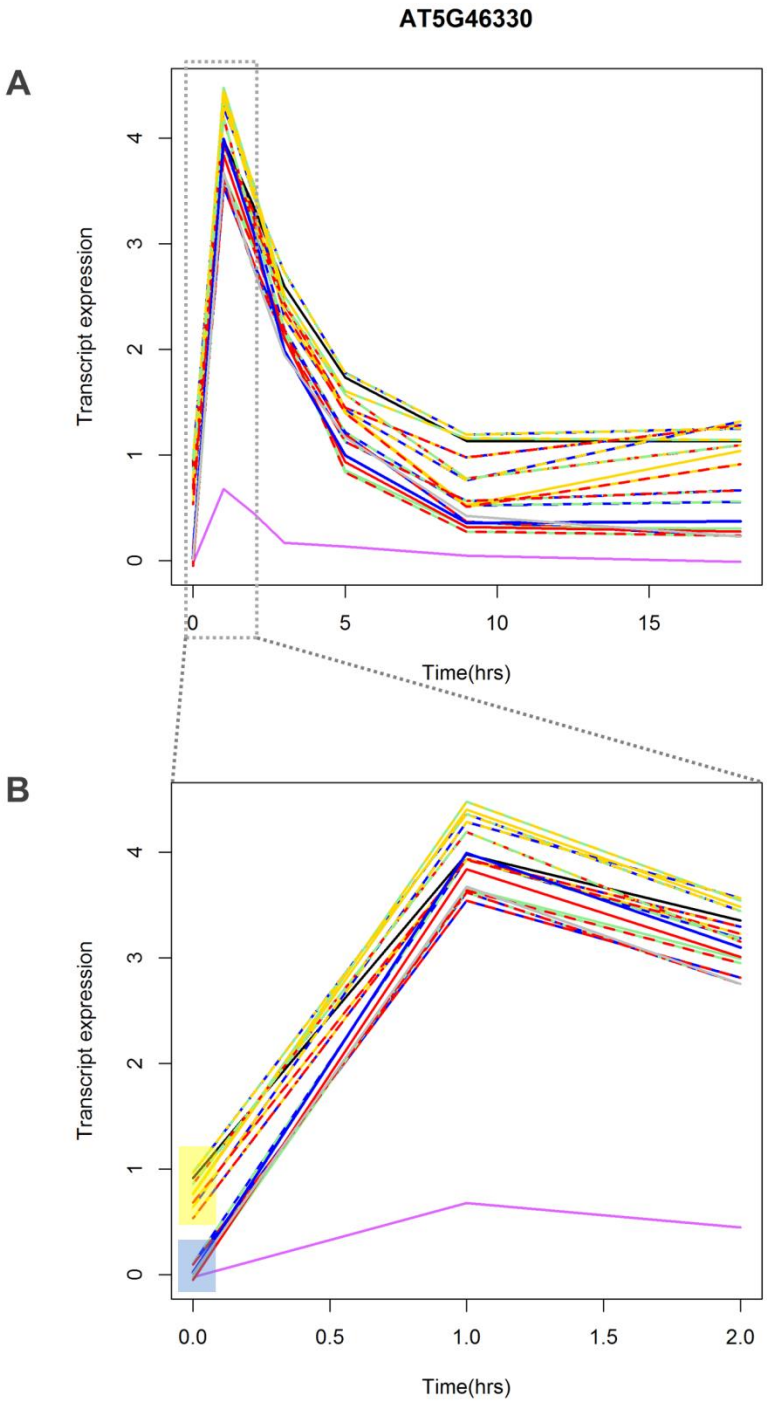
S5 Fig. Signaling allocations for alternative PAD4 sector activity markers. Only signaling allocation contributions with q-values < 0.05 (Storey's FDR) are shown. The allocations for each gene were scaled for visualization using the extreme value of each. Note that no or very small signaling allocation contributions are found where PAD4 or P is not included in a contribution's label, which was one of the criteria we used to select PAD4 sector activity marker genes.



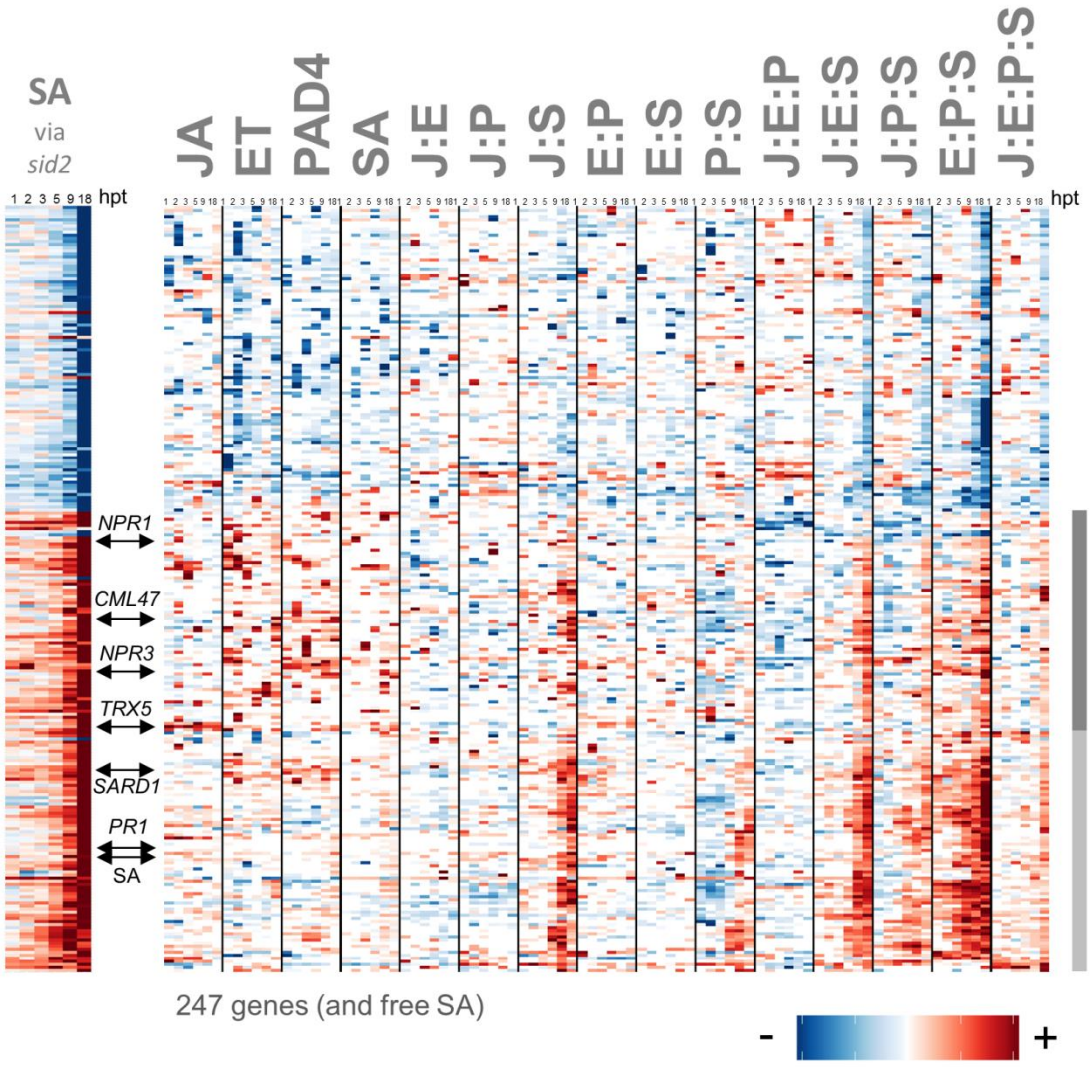
S6 Fig. A time course of the log_2 free SA hormone concentrations per genotype. (A) Mean log_2 SA hormone concentrations and (B) mean log_2 SA concentrations relative to each genotype's constitutive SA level. The mean value across biological replicates is shown for each genotype at each time point. Colors indicate genotype by presence, not absence of a sector: red (JA), yellow (ET), green (PAD4), blue (SA). Other genotypes: black (wild type), gray (quadruple mutant), purple (*fls2*). The red arrow in (B) indicates the *dde2* genotype. Data were collected at 0, 1, 2, 3, 5, 9, and 18 hours after flg22 treatment.



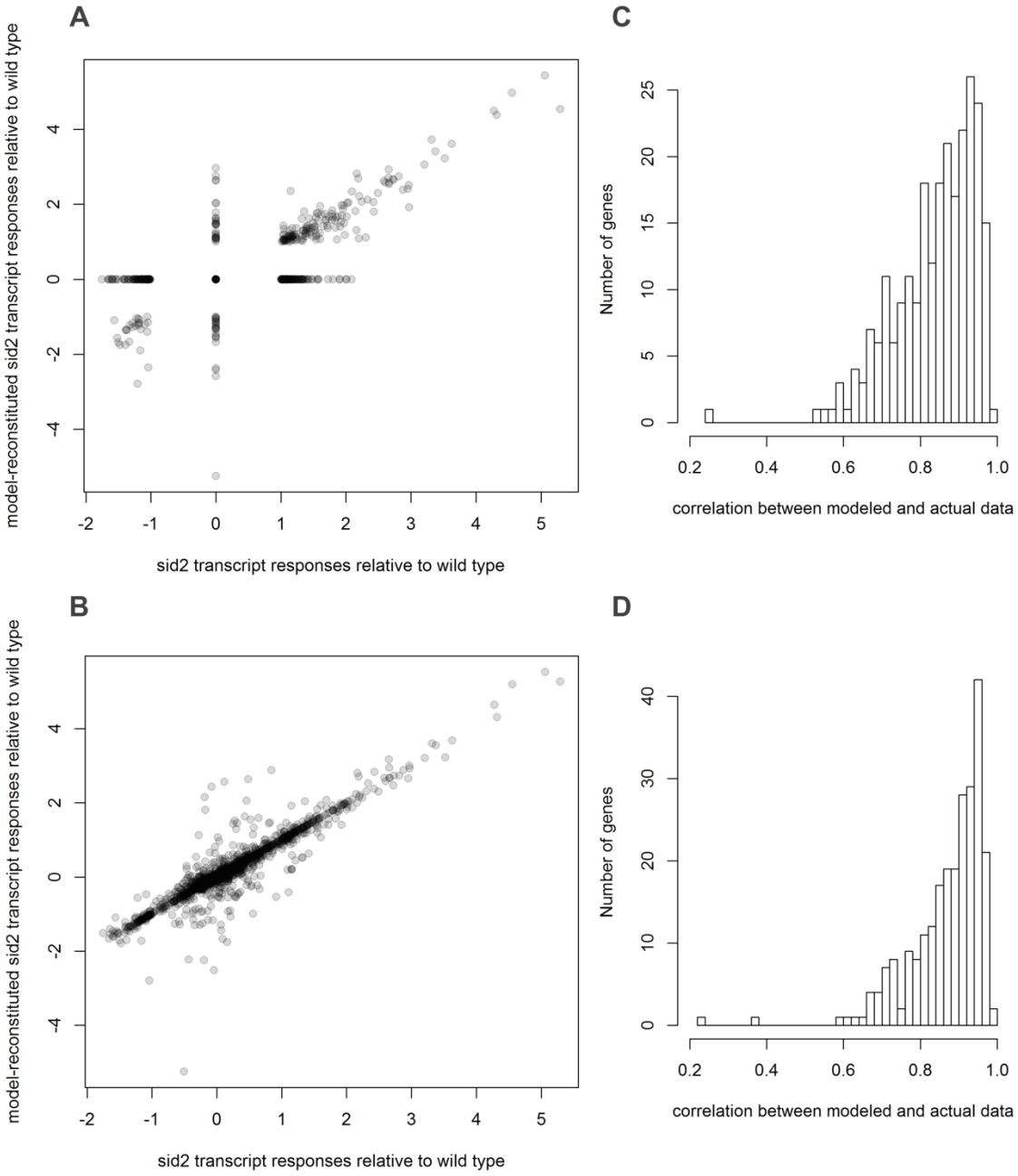
S7 Fig. A time course of the adjusted log_2 free JA hormone concentrations per genotype. Mean adjusted log_2 JA levels across biological replicates are shown for all genotypes and time points profiled. Colors indicate genotype by presence, not absence of a sector: red (JA), yellow (ET), green (PAD4), blue (SA). Other genotypes: black (wild type), gray (quadruple mutant), purple (*fls2*). Data were collected at 0, 1, 2, 3, 5, 9, and 18 hours after flg22 treatment.



S8 Fig. *FLS2* transcript levels in the *ein2*-containing genotypes. *FLS2* transcript levels in the *ein2*-containing genotypes are increased to nearly wild type levels within 1 hour after flg22 treatment. (A and B) Mean expression of AT5G46330 (*FLS2*) across biological replicates, for all genotypes and time points profiled. Colors indicate genotype by presence, not absence of a sector: red (JA), yellow (ET), green (PAD4), blue (SA). Other genotypes: black (wild type), gray (quadruple mutant), purple (*fls2*). (B) A zoom-in on early time points. Yellow box: *EIN2*-containing genotypes; blue box: *ein2*-containing genotypes and *fls2*. Data were collected at 0, 1, 2, 3, 5, 9, and 18 hours after flg22 treatment.

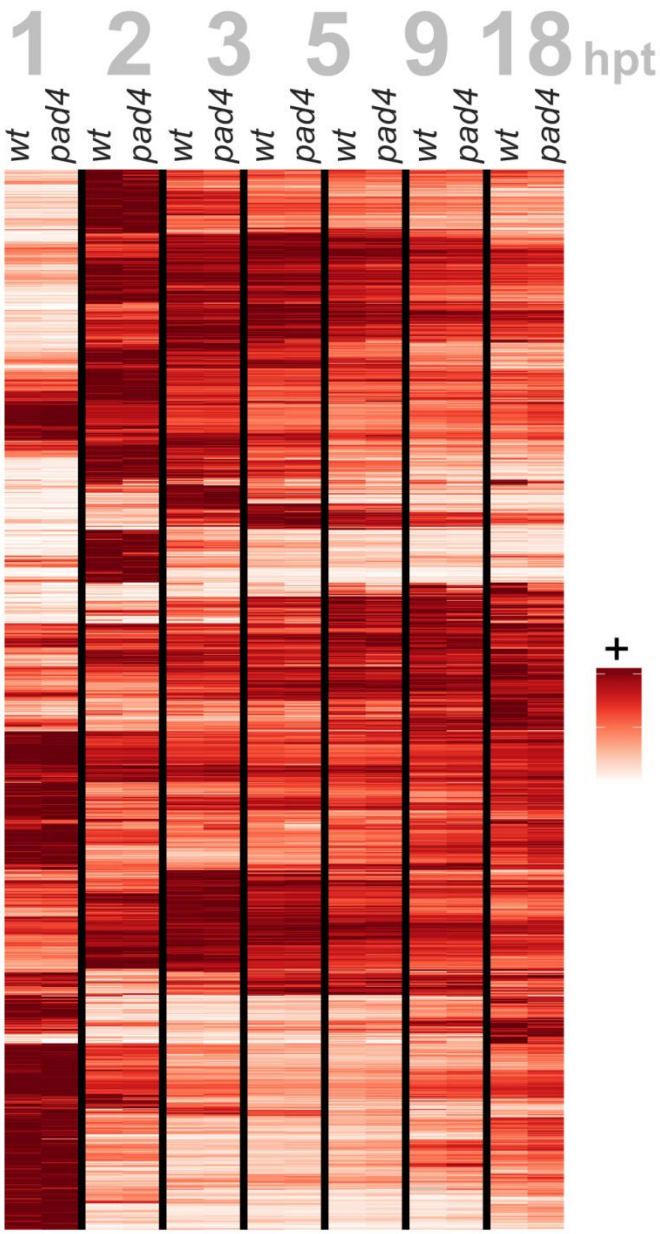


S9 Fig. Figure 4-6 without q-value based regularization, and without masking insignificant changes in *sid2*. As in Figure 4-6, two distinct signaling allocation signatures are seen (light gray and dark gray bars alongside the clustergram) for traditionally-defined SA-dependent genes. This traditional definition of SA-dependence is based on gene transcript responses in the *sid2* single-sector mutant relative to wild type responses. The row ordering of genes and hormones here is the same as in Figure 4-6.

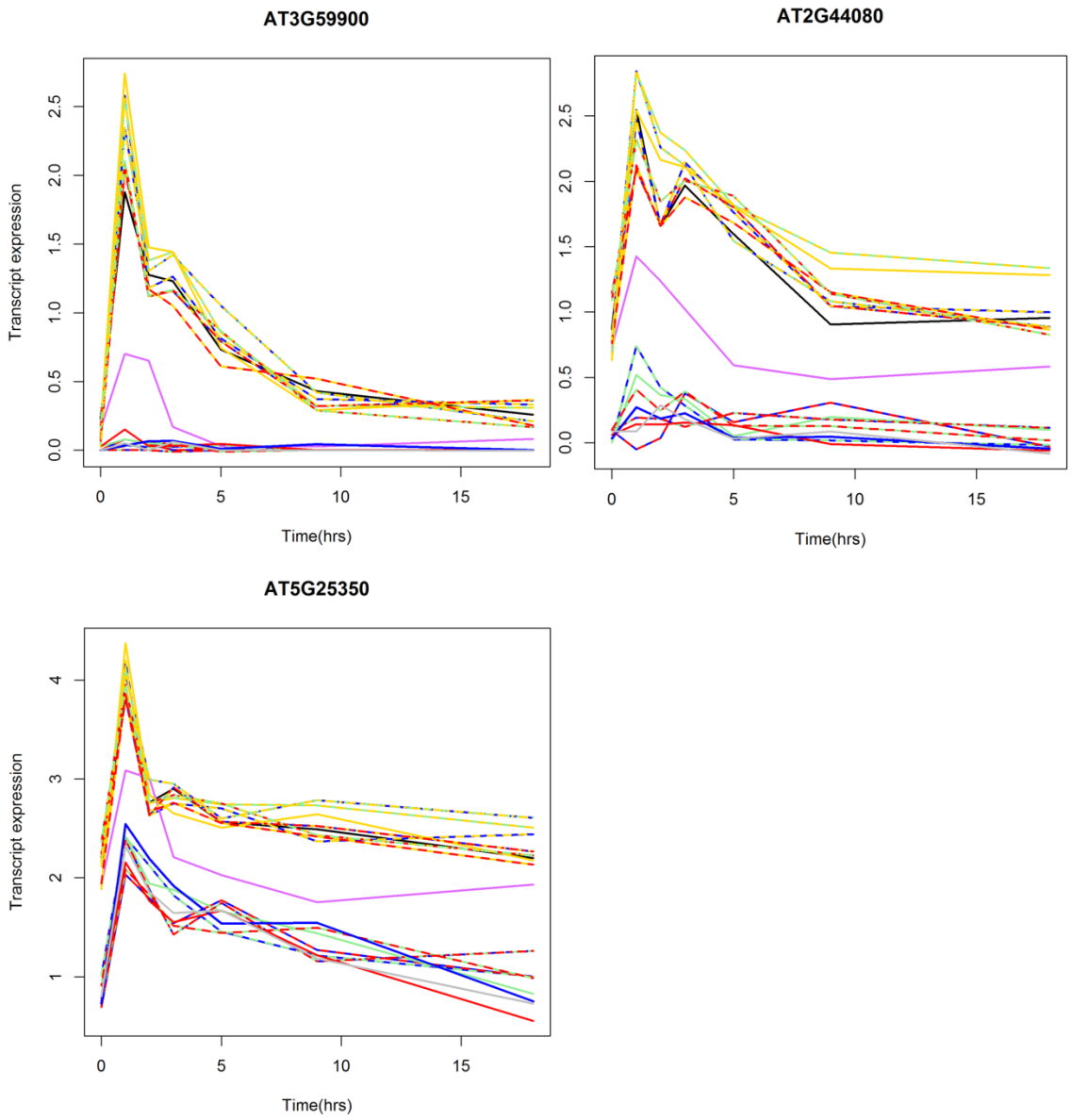


S10 Fig. Model correlations with *sid2* transcript response changes relative to wild type for traditionally-defined SA-dependent genes. (A) Correlation between modeled and actual *sid2* transcript changes relative to wild type for the 247 genes and SA in Figure 4-6: Pearson correlation 0.71. (B) Correlation between modeled and actual *sid2* transcript responses relative to wild type responses for the same genes and SA in S9 Fig: Pearson correlation 0.92. (C) and (D) The median Pearson correlation of models with data across all genotypes and time points for the Figure 4-6 genes and SA was 0.86 after q-value based regularization and 0.89 before q-value based regularization. (C) and (D) show the distributions of Pearson correlations of model-reconstituted transcript expression and hormone abundance levels with actual expression or hormone levels for (C) the signaling allocations (i.e. models) in Figure 4-6 and (D) the signaling allocations (i.e. models) in S9 Fig. For all Figure 4-1 genes, across all genotypes and time points,

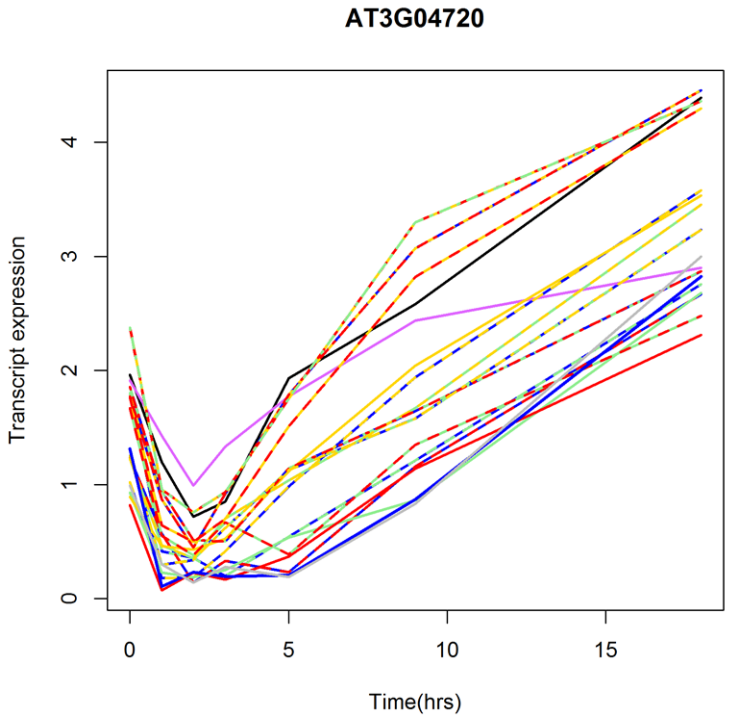
median Pearson correlations were 0.84 and 0.90, after and before, respectively. (See S24 Fig and S25 Fig for histograms analogous to (C) and (D) for all Figure 4-1 genes.)



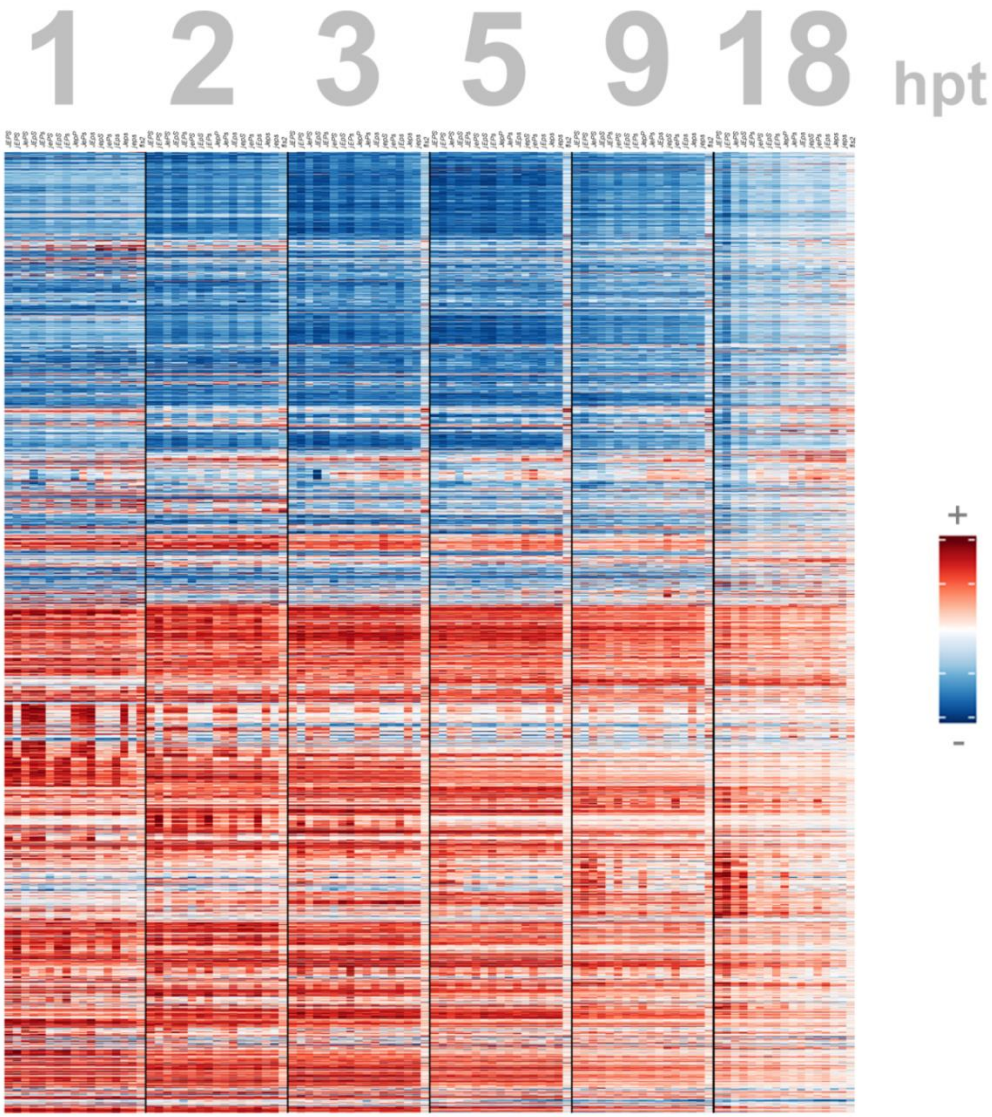
S11 Fig. Few genes show differential expression between wild type and *pad4* at any sampled time point. Mean transcript expression and hormone abundance levels at each sampled time point for the 3211 genes and 2 hormones in Figure 1. There is a trend for induced genes to be slightly less induced in *pad4* (*JEpS*) than in wild type (*JEPS*), and for repressed genes to be slightly less repressed, but these effects are small. Rows have been re-ordered via clustering of the expression patterns shown; rows have been individually scaled for better visualization using the extreme value of each row.



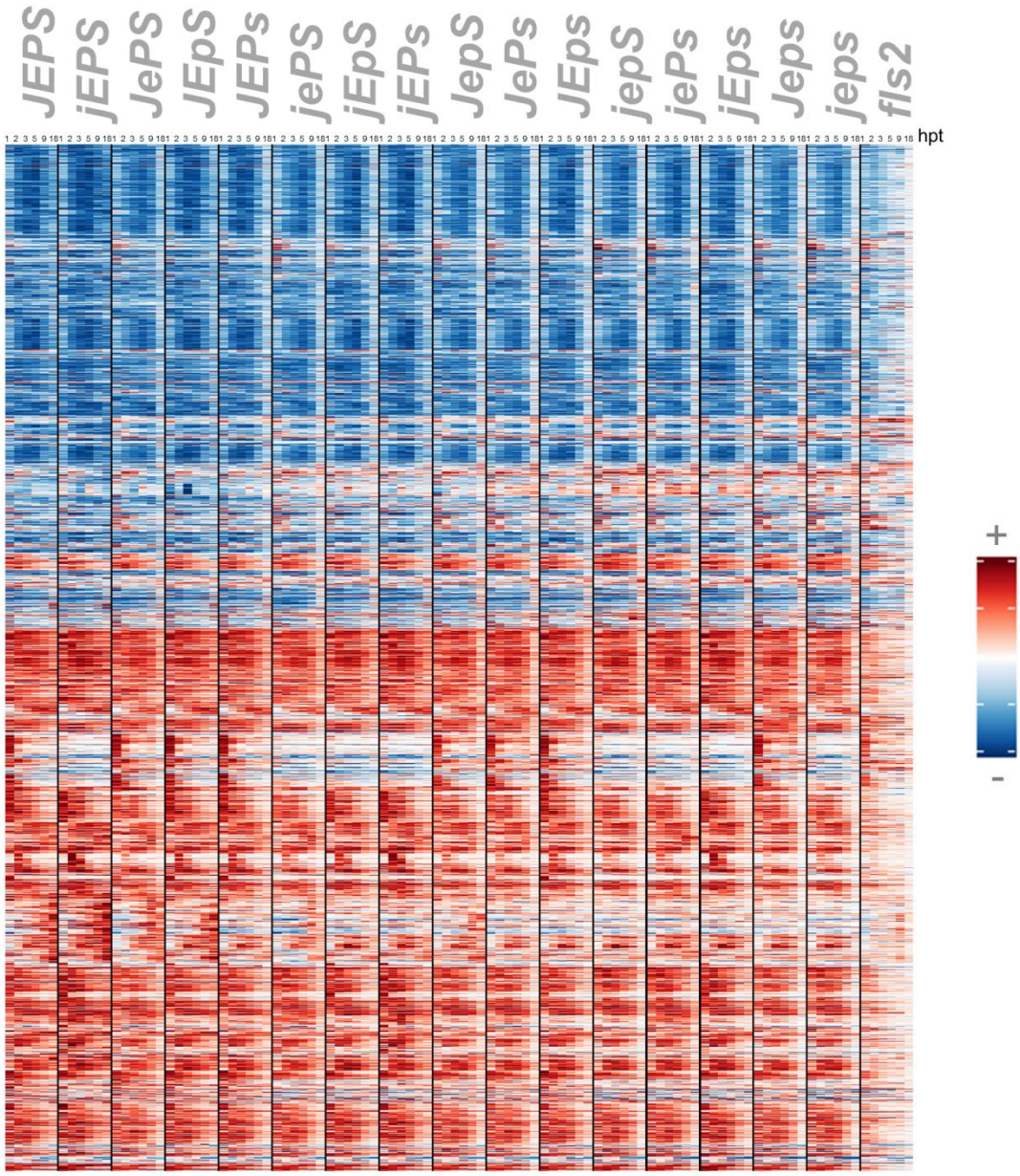
S12 Fig. An early ethylene spike is evident in ET sector activity marker genes. Mean transcript expression levels of ET sector activity marker genes, and the ACC-inducible control EBF2, across all genotypes and time points profiled. (A) *ARGOS* (AT3G59900), (B) *ARL* (AT2G44080), (C) *EBF2* (AT5G25350). Colors indicate genotype by presence, not absence, of a sector: red (JA), yellow (ET), green (PAD4), blue (SA). Other genotypes: black (wild type), gray (quadruple mutant), purple (*fls2*). Data were collected at 0, 1, 2, 3, 5, 9, and 18 hours after flg22 treatment.



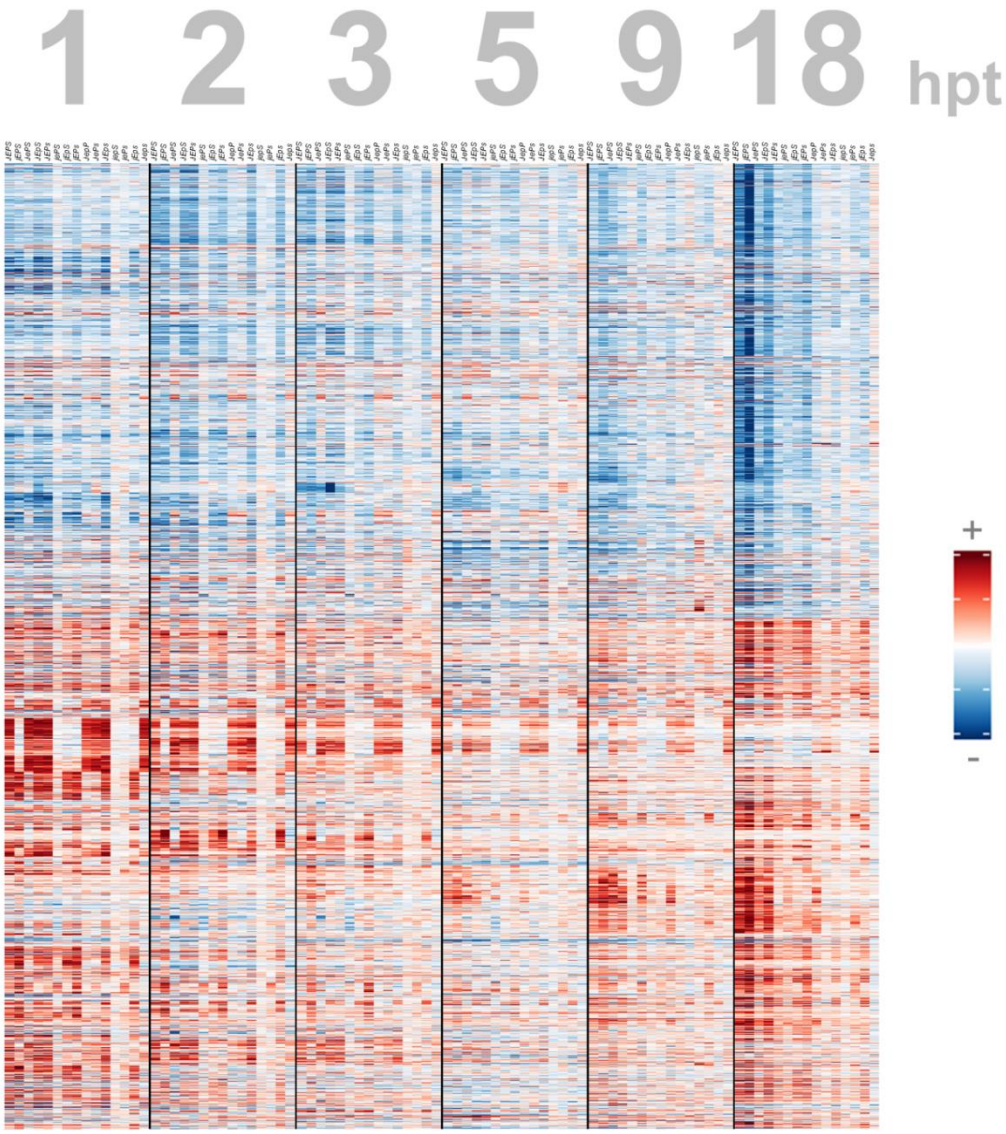
S13 Fig. *PR4* was down-regulated at early time points after flg22 treatment.
PR4 was down-regulated at early time points but recovered by 9 hrs after flg22 treatment, and was induced at 18 hrs. Mean transcript expression levels are plotted across all genotypes and time points profiled. Colors indicate genotype by presence, not absence, of a sector: red (JA), yellow (ET), green (PAD4), blue (SA). Other genotypes: black (wild type), gray (quadruple mutant), purple (*fls2*). Data were collected at 0, 1, 2, 3, 5, 9, and 18 hours after flg22 treatment.



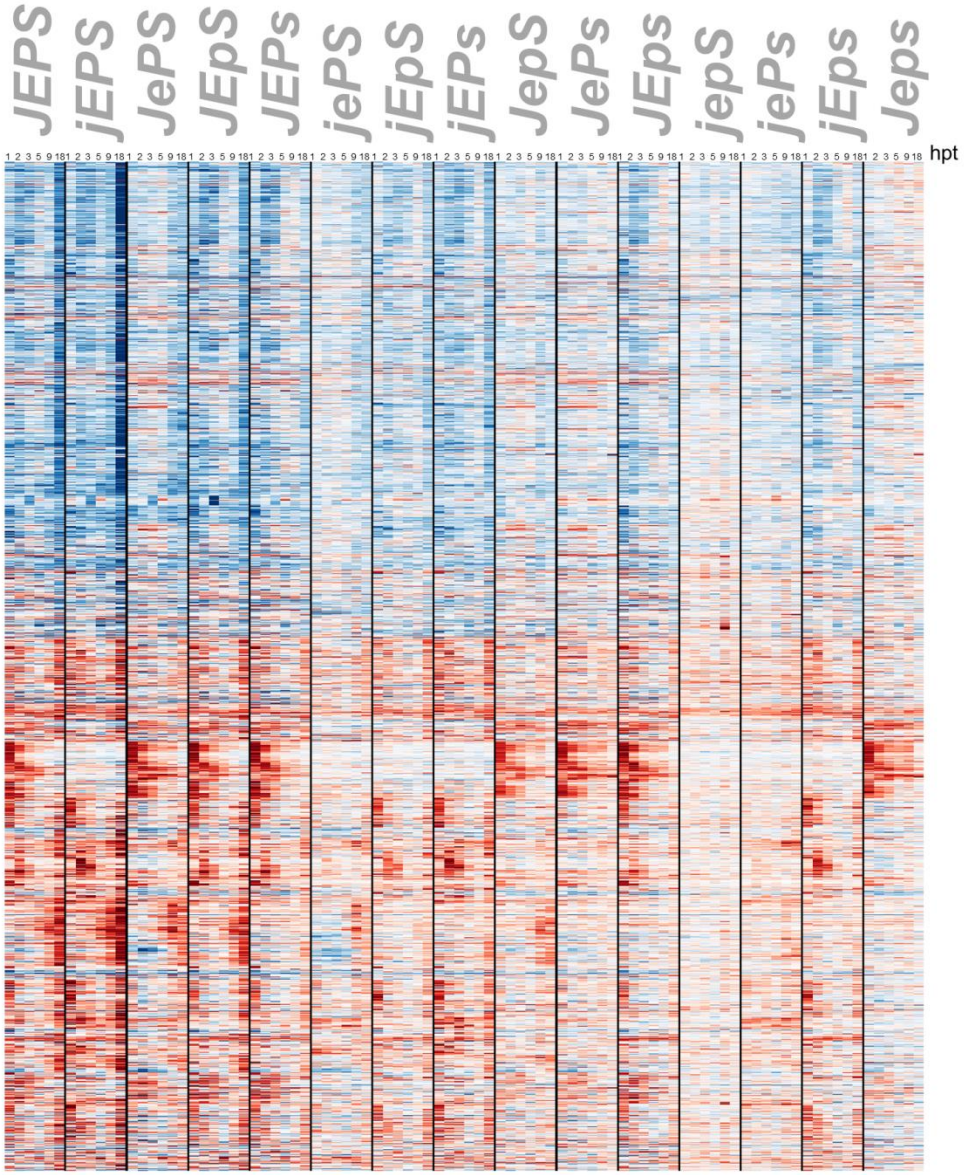
S14 Fig. Transcript expression level responses for the genes and hormones in Figure 4-1. A clustergram is shown for the transcript expression (or hormone) levels relative to 0 hr levels for the 3211 genes and 2 hormones in Figure 4-1. A transcript response is the expression level relative to the 0 hr expression level of a transcript. For most of these genes, the transcript response in the quadruple mutant remains highly dynamic; in fact while the JA-ET-PAD4-SA network significantly contributes to the transcript responses of each of these genes and hormones, the qualitative dynamic responses across genotypes remain intact. It is surprising that mutations affecting flg22-PTI do little to qualitatively alter the dynamics of the flg22-inducible transcriptional response, but rather mainly affect the strength of the response at each time point. Transcript (or abundance) responses for each gene (or hormone) were individually scaled for visualization using the extreme value of each row. Combinatorial sector genotypes are labeled by the presence (*J*,*E*,*P*,*S*) or absence (*j*,*e*,*p*,*s*) of each sector, for example the *dde2 ein2* genotype is written as '*jePS*'. Each major column contains the genotypes *JEPS*, *jEPS*, *JePS*, *JEpS*, *JEPs*, *jePS*, *jEpS*, *jEPs*, *JepS*, *JePs*, *JEps*, *jepS*, *jePs*, *jEps*, *Jeps*, *jeps*, and *fls2*, in that order. Transcript responses are shown for 1, 2, 3, 5, 9, 18 hours post treatment (hpt) with flg22. The row ordering of genes and hormones here is the same as in Figure 4-1.



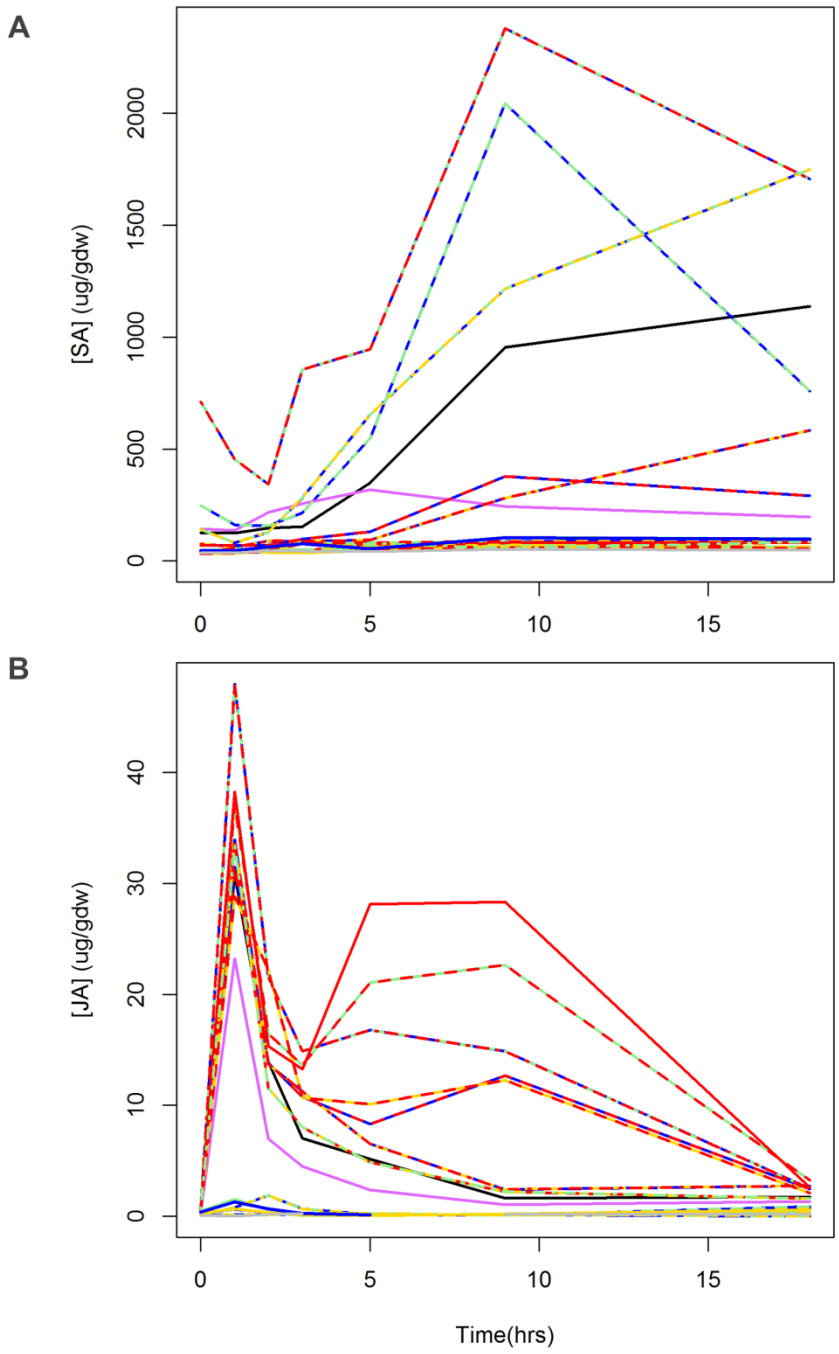
S15 Fig. Transcript expression dynamics for most genes are very similar across the combinatorial sector genotypes. The data in this figure is identical to the data in S14 Fig; the columns have merely been re-ordered. Transcript responses are shown for the 3211 genes and 2 hormones in Figure 4-1. A transcript response is the expression level relative to the 0 hr expression level of a transcript. Transcript (or abundance) responses for each gene (or hormone) were individually scaled for visualization using the extreme value of each row. Combinatorial sector genotypes are labeled by the presence (*J*,*E*,*P*,*S*) or absence (*j*,*e*,*p*,*s*) of each sector, for example the *dde2 ein2* genotype is written as '*jePS'*. Transcript responses are shown for 1, 2, 3, 5, 9,18 hours post treatment (hpt) with flg22. The row ordering of genes and hormones here is the same as in Figure 4-1.



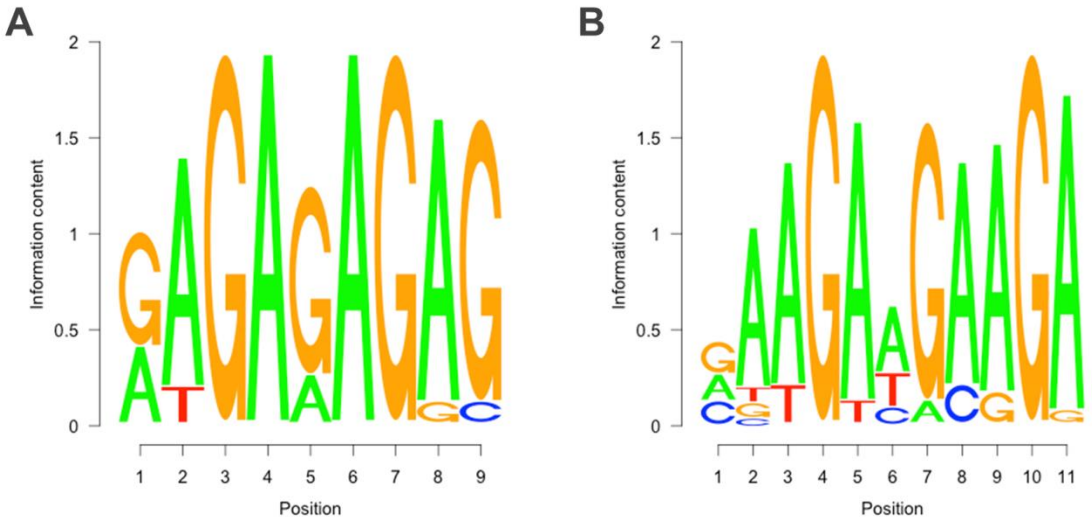
S16 Fig. Genotype differences in transcript expression dynamics are evident in flg22-induced transcript response changes. A clustergram is shown for the transcript response changes of the genes and hormones in Figure 4-1. Transcript response changes (transcript expression relative to 0 hr expression in a combinatorial mutant – transcript expression relative to 0 hr expression in the quadruple mutant) for the 3211 genes and 2 hormones in Figure 4-1 are shown, ordered by time point. The values in this figure were calculated by subtracting the quadruple mutant transcript expression responses in S14 Fig from those of the other combinatorial genotypes in S14 Fig. Transcript (or abundance) response changes for each gene (or hormone) were individually scaled for visualization using the extreme value of each row. Combinatorial sector genotypes are labeled by the presence (*J*,*E*,*P*,*S*) or absence (*j*,*e*,*p*,*s*) of each sector, for example the *dde2 ein2* genotype is written as '*jePS*'. Each major column contains the genotypes *JEPS*, *jEPS*, *JePS*, *JEpS*, *JEPs*, *jePS*, *jEpS*, *jEPs*, *JepS*, *JePs*, *JEps*, *jepS*, *jePs*, *jEps*, and *Jeps*, in that order. Transcript response changes are shown for 1, 2, 3, 5, 9,18 hours post treatment (hpt) with flg22. The row ordering of genes and hormones here is the same as in Figure 4-1.



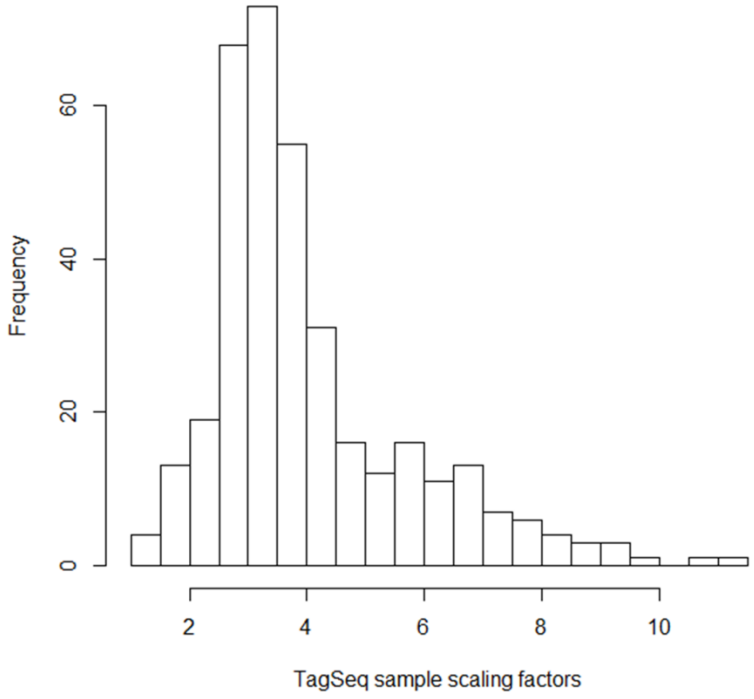
S17 Fig. Genotype differences in transcript expression dynamics are evident in transcript response changes. The data in this figure is identical to that in S16 Fig; the columns have merely been re-ordered. Transcript response changes (transcript expression relative to 0 hr expression in a combinatorial mutant – transcript expression relative to 0 hr expression in the quadruple mutant) for the 3211 genes and 2 hormones in Figure 4-1 are shown, ordered by genotype. The values in this figure were calculated by subtracting the quadruple mutant transcript expression responses in S15 Fig from those of the other combinatorial genotypes in S15 Fig. Transcript (or abundance) response changes for each gene (or hormone) were individually scaled for visualization using the extreme value of each row. Combinatorial sector genotypes are labeled by the presence (*J*,*E*,*P*,*S*) or absence (*j*,*e*,*p*,*s*) of each sector, for example the *dde2 ein2* genotype is written as '*jePS'*. Transcript response changes are shown for 1, 2, 3, 5, 9,18 hours post treatment (hpt) with flg22. The row ordering of genes and hormones here is the same as in Figure 4-1.



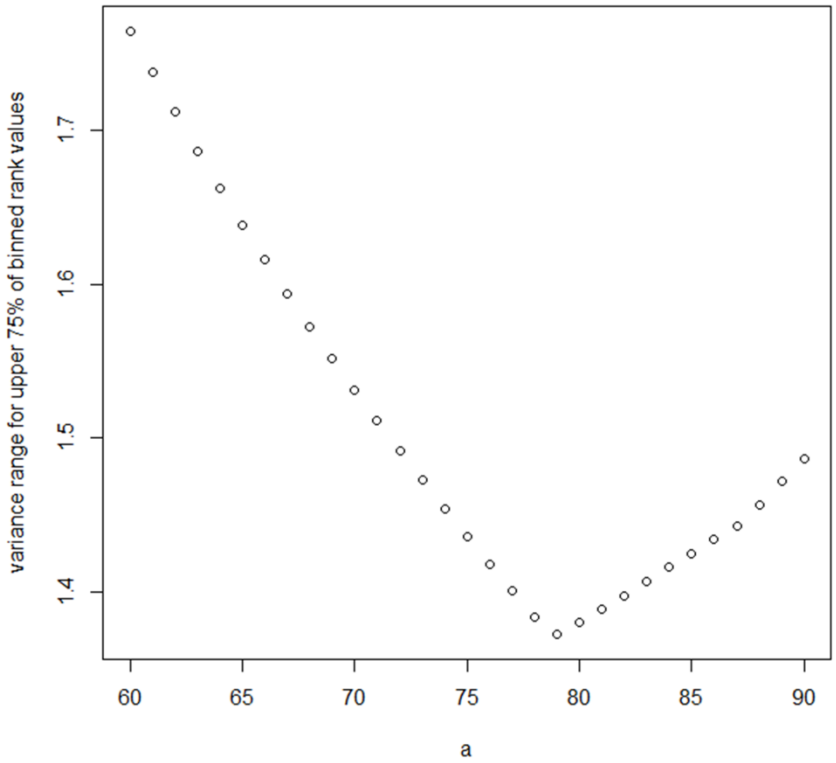
S18 Fig. Time courses of free SA and free JA hormone concentrations per genotype. (A) Free SA hormone concentrations and (B) free JA hormone concentrations in units of micrograms (of hormone) per gram dry weight (of plant tissue). Replicate means were calculated from $\log_2$ values across all genotypes and time points profiled. For JA data, the two outlier points were masked for this mean calculation. Colors indicate genotype by presence, not absence of a sector: red (JA), yellow (ET), green (PAD4), blue (SA). Other genotypes: black (wild type), gray (quadruple mutant), purple (*fls2*). Data were collected at 0, 1, 2, 3, 5, 9, and 18 hours after flg22 treatment.



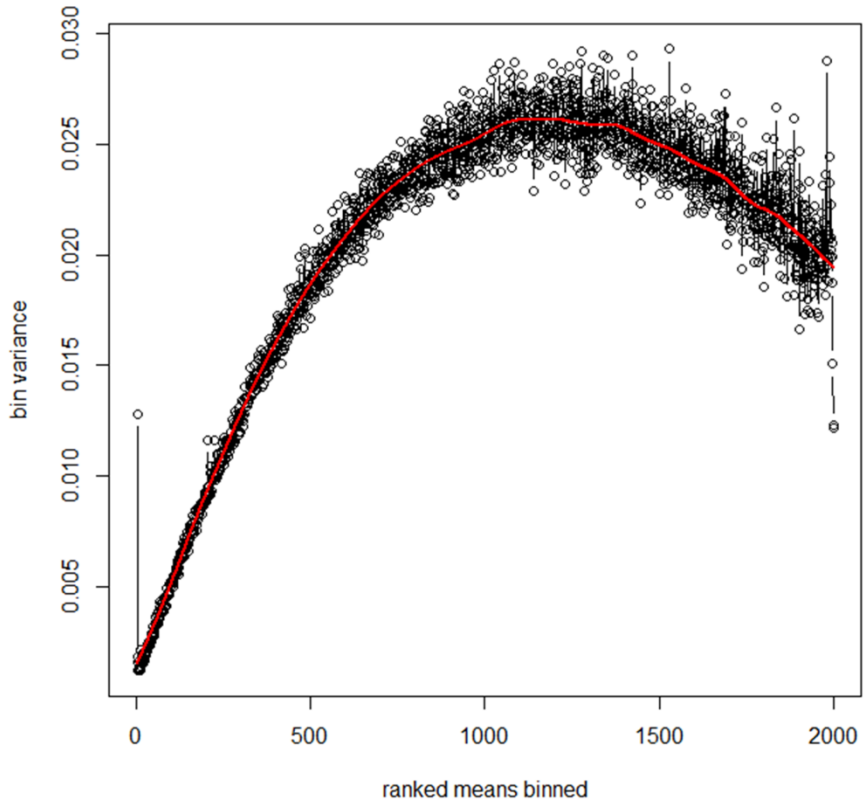
S19 Fig. Representative GA-rich and AAG-rich motifs found in the promoters of flg22-responsive, network-responsive genes. Genes with similar signaling allocation signatures (by cosine distance complete-linkage agglomerative hierarchical clustering) were subjected to MEME-based de novo motif discovery. Nearly every cluster tested of 57 total (similarity cutoff height 1.2; clusters with more than 20 genes were fed to MEME) showed evidence of GA-rich and AAG-rich motifs. (A) An example GA-rich motif from cluster 11. (B) An example AAG-rich motifs from cluster 1. (See S5 Data for the full set of motifs discovered for all clusters.) Motif logos were created using the Bioconductor R package 'seqLogo'.



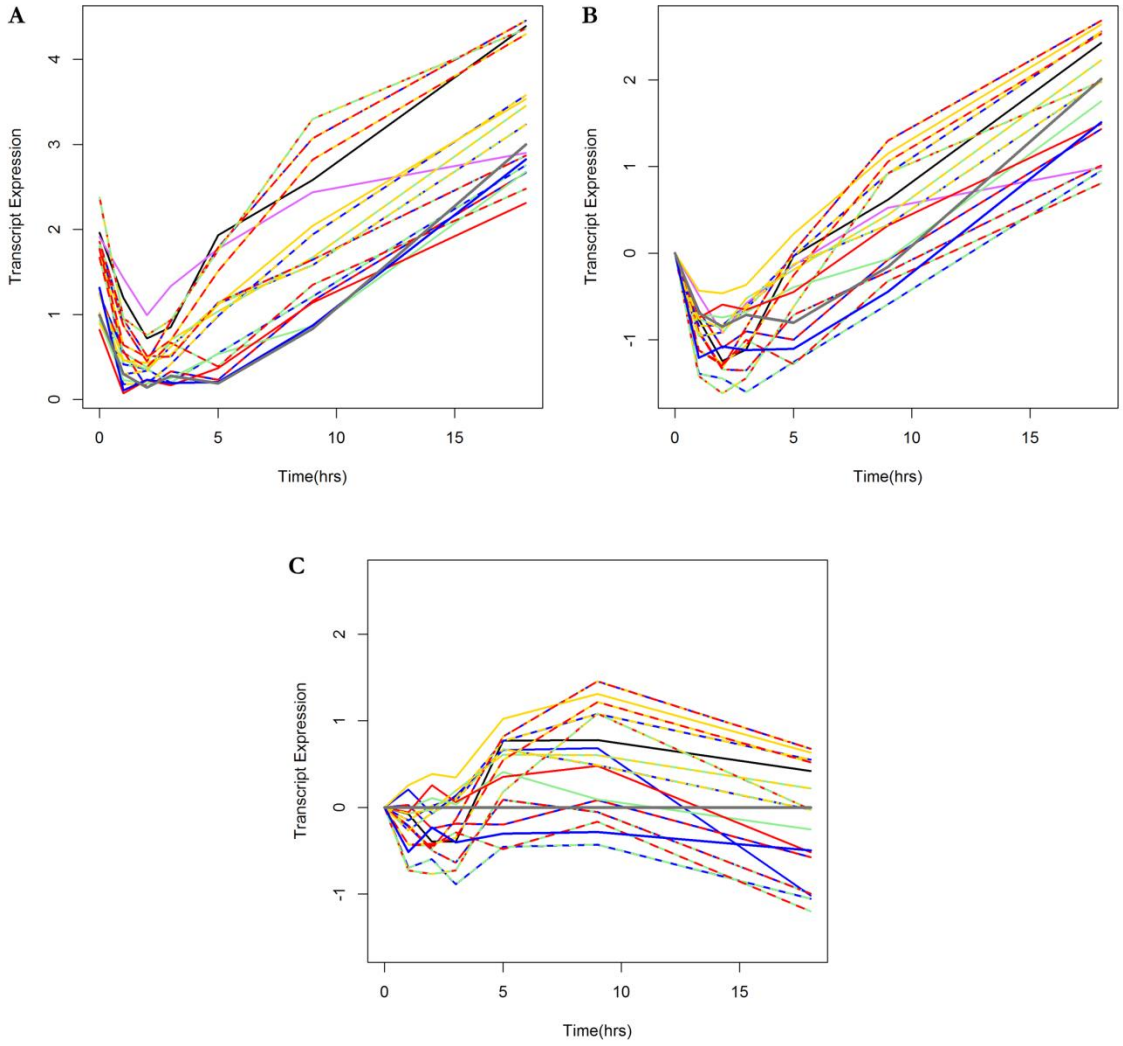
S20 Fig. Histogram of sample scaling factors for all 357 Tag-Seq datasets.



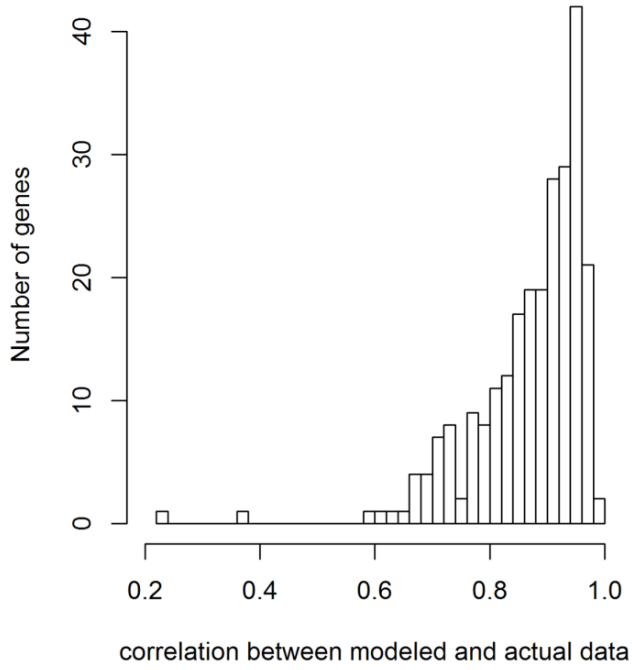
S21 Fig. Mean residual variance for the searched range of a parameters. Residual variance for the upper 75% of binned mean values for the searched range of a parameters. a=79 was selected.



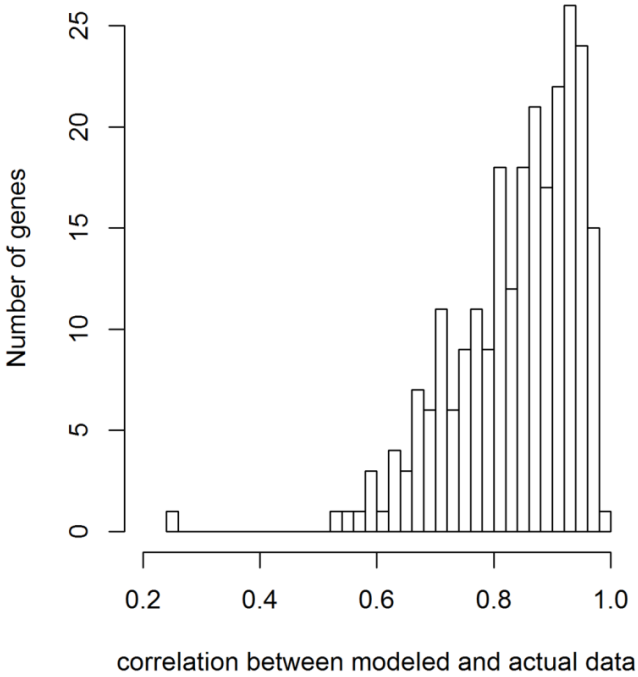
S22 Fig. Residual variance for mean-ranked Tag-Seq data. Fitted ranked mean values were binned into 2000 bins; variance in associated residuals was calculated for each bin. Note that the variance is fairly stable for the upper 75% of bins. The red line is a loess fit to the bin variance values.



S23 Fig. Calculation of transcript response change values for the gene PR4.
(A) Mean transcript expression level values. (B) Mean transcript expression response values; these values were calculated by subtracting the 0 hr values from each genotype's time course. (C) Mean transcript response change values. Time course values of the quadruple mutant (*jeps*) have been subtracted from the 0 hr adjusted time courses of the other combinatorial mutants. Colors indicate genotype by presence, not absence of a sector: red (JA), yellow (ET), green (PAD4), blue (SA). Other genotypes: black (wild type), gray (quadruple mutant), purple (*fls2*). *Fls2* is not included in (C) since *fls2* is a good control for wild type, but not for any of the other mutants. Profiled time points: 0, 1, 2, 3, 5, 9, 18 hrs after flg22 treatment.



S24 Fig. Correlation between model-predicted data and actual data. For the 3211 genes and 2 hormones in Figure 4-1: Pearson correlation between expression data and predicted expression data for the lasso- and AICc-regularized signaling allocation models. Median correlation: 0.89.



S25 Fig. Correlation between model-predicted data and actual data. For the 3211 genes and 2 hormones in Figure 4-1: Pearson correlation between expression data and predicted expression data for the lasso- and AICc-regularized signaling allocation models which were further regularized via contribution coefficient q-values. Median correlation: 0.84.

AGI locus identifier	Gene Name
AT4G31800	WRKY18
AT4G29810	MKK2
AT1G73805	SARD1
AT3G56710	SIB1
AT2G14560	LURP1
AT2G40750	WRKY54
AT3G56400	WRKY70
AT5G54610	BDA1
AT1G03850	GRXS13
AT4G14400	ACD6
AT2G46440	CNGC11
AT5G24530	DMR4
AT3G11340	UGT76B1
AT1G02450	NIMIN1
AT5G22570	WRKY38
AT5G01900	WRKY62
AT1G64280	NPR1
AT4G23810	WRKY53
AT5G08790	NAC
AT2G13790	BAK1
AT1G51660	MKK4
AT1G11310	MLO2
AT5G45110	NPR3
AT1G45145	TRX5

S1 Table. Figure 4-6 cluster C and cluster D genes include genes with SA-related functions. List of traditionally-defined SA-dependent genes in Figure 4-6 that also have literature reports of immune-related phenotypes for mutant allele(s) of those genes. Gene sets from both Figure 4-6 cluster C (dark gray) and Figure 4-6 cluster D (light gray) both contain SA-related genes. *PAD4* and *SID2* are in Figure 4-6 cluster C (dark gray cluster), but these genes are omitted as they were directly mutated in our dataset. Literature mutant phenotypes were identified via curation of NCBI GeneRIF annotation (ThaleMine v1.8.0, Araport 11 Pre-release 3) (Krishnakumar et al. 2015)**.**

4.6.4 Additional Supporting Information

S1 Data. Signaling allocations. (available upon request)

S2 Data. Signaling allocation contribution p-values. (available upon request)

S3 Data. Raw hormone data. (available upon request)

S4 Data. Pre-processed hormone data. (available upon request)

S5 Data. Motif position weight matrices and database matches. Position weight matrices (PWMs) for discovered motifs. (available upon request) Motifs without

database matches are labeled with ‘denovo’. Motifs with matches to the CisBP database are labeled ‘matched’. Columns are in the standard PWM order (A, C, G, T); rows correspond to motif positions. For each matched motif, there is a corresponding file with a list of CisBP motif names and associated transcription factor (TF) AGI locus identifiers.

S6 Data. Cluster gene lists and signaling allocation clustergrams. (available upon request) Clusters of signaling allocations used as input for motif discovery (see S5 Data for discovered motifs).

S3 Text. Infiltration Schedules. (available upon request)

S4 Text. AGI locus identifiers for Figure 4-1 genes. (available upon request)

S5 Text. AGI locus identifiers for Figure 4-1 column a genes. (available upon request)

S6 Text. AGI locus identifiers for Figure 4-1 column b genes. (available upon request)

S7 Text. AGI locus identifiers for Figure 4-5A genes. (available upon request)

S8 Text. AGI locus identifiers for Figure 4-5B genes. (available upon request)

S9 Text. AGI locus identifiers for Figure 4-5C genes. (available upon request)

S10 Text. AGI locus identifiers for Figure 4-6 cluster A genes. (available upon request)

S11 Text. AGI locus identifiers for Figure 4-6 cluster B genes. (available upon request)

S12 Text. AGI locus identifiers for Figure 4-6 cluster C genes. (available upon request)

S13 Text. AGI locus identifiers for Figure 4-6 cluster D genes. (available upon request)

S14 Text. AGI locus identifiers for Figure 4-7A genes. (available upon request)

S15 Text. AGI locus identifiers for Figure 4-7B genes. (available upon request)

S16 Text. AGI locus identifiers for Figure 4-7C genes. (available upon request)

S17 Text. AGI locus identifiers for overlapping genes between Figure 4-6 cluster C and Figure 4-7A. (available upon request)

S18 Text. AGI locus identifiers for overlapping genes between Figure 4-7B cluster 1 and Figure 4-7A. (available upon request)

S19 Text. AGI locus identifiers for Figure 4-8 genes. (available upon request)

S20 Text. AGI locus identifiers for Figure 4-9A genes. (available upon request)

S21 Text. Hormones evaluated for relative responsive abundance relative to quad. (available upon request)

S22 Text. Enriched GOannotations for Figure 4-6 cluster A genes. (available upon request)

S23 Text. Enriched GOannotations for Figure 4-6 cluster B genes. (available upon request)

S24 Text. Enriched GOannotations for Figure 4-6 cluster C genes. (available upon request)

S25 Text. Enriched GOannotations for Figure 4-6 cluster D genes. (available upon request)

S2 Table. Comparison of enriched GO annotations for Figure 4-6 clusters C and D. (available upon request)

Chapter 5.
Conclusion

In this dissertation, I studied the mechanisms by which the JA-ET-PAD4-SA network is activated, regulates itself, and regulates the plant transcriptome after plant immune stimulus with bacterial flagellin (flg22). Mathematical descriptions of plant immune signaling mechanisms are largely lacking in the literature; the work described in this dissertation extends the work of the few available prior studies (Tsuda et al. 2009; Kim et al. 2014). In this dissertation, I searched for mechanistic models that can explain network dynamics (Chapter 3), and I quantified mechanisms by which the JA-ET-PAD4-SA network dynamically regulates the flg22-responsive plant transcriptome (Chapter 4).

5.1 Major Findings

In Chapter 3, I described delay differential equations (DDEs) that show promise of having the needed form and flexibility to capture network sector dynamics after flg22 treatment. These DDE models followed time-interpolated sector activity data well, but there are approximately as many model parameters as there are mean data values underlying the time-interpolations. Thus, more data will be required to establish time-interpolations that truly follow sector activity changes, in order to validate that these DDE models are capturing network mechanisms and not interpolation artifacts.

In Chapter 4, I adapted the signaling allocation models in (Tsuda et al. 2009), together with the model regularization techniques used in (Kim et al. 2014) into a hybrid modeling approach. I fit regularized signaling allocation models to dynamic transcriptome and hormone data collected after flg22 treatment. I described quantitative regulatory mechanisms for how the JA-ET-PAD4-SA network regulates the flg22-responsive transcriptome, as well as the hormones JA and SA. Notably, I documented regulatory mechanisms that underlie an important emergent property of the plant immune signaling network, network buffering.

5.2 Future Directions

The work in this dissertation leads to several interesting follow-up directions. The DDE models of PAD4 activity in Chapter 3 predict several intermediate nodes where signals are expected to converge. A straightforward way to test these predictions is to mine our RNA-Seq transcriptome data for genes that follow the predicted activity of these nodes. When more than two signals converge on a node, it is unlikely that signal convergence takes place in a single step. It may be possible to discover, from our RNA-Seq data, in what order signal convergence occurs for each step, if transcriptional proxies can be found for some signal combinations but not others at a given predicted node.

Another important next step is to find DDE models for all four sectors, not just PAD4, and to combine these models into a single set of coupled DDE equations. It is worth noting that the inter-genotype differences in sector dynamics for the other three sectors—JA, ET and SA—are less distinguishable than for PAD4. The diverse PAD4 dynamics in different genotypes provided strong modeling constraints for the PAD4 DDE model; for JA, ET and SA, there will be fewer constraints. As a result, there may be multiple different individual DDE models that can explain each of these sectors. However, when the four sectors are combined together into a single set of coupled DDE equations, dynamic constraints across the network may be useful in selecting the best-performing model of these potential models for each of the JA, ET, and SA sectors. For the full combined DDE network model, modeled values for each sector need to be able to predict modeled values for other sectors given only network inputs. Finding such a combined model will be a crucial modeling milestone; such a model can make type II predictions that can be tested. Notably, such a model can ask and answer the question, "How would the network behave with a different set of input signals?" The Kim model (Kim et al. 2014) suggests two biologically relevant network input signals that are different from those of flg22, and which we could use to make and test these type II predictions: the PTI network inputs from the MAMPs elf18 and chitosan.

The work in Chapter 4 also opened up interesting avenues for future discovery. The most unexpected result was the complete buffering of the *pad4-1* mutation's effect on the flg22 transcriptional response by the rest of the network. This is surprising because

pad4-1 has a strong flg22-PTI phenotype against *P. syringae* pv. *tomato* DC3000; how can the flg22 transcriptome response show virtually no effects? We offered the hypothesis that perhaps flg22 activates a mediator of the PAD4 sector's effect on immunity, proposing EDS1 as this potential activator. However, testing this hypothesis is not straightforward. Whether EDS1 is required for PAD4's function in immunity could be tested by profiling an *eds1* mutant, a *pad4* mutant, and an *eds1 pad4* double mutant after pathogen treatment. Note that flg22 treatment will not work; our hypothesis is that flg22 merely activates an immunity mediator, but that the effects of this mediator are not seen without pathogen treatment. If these three mutant profiles are very similar, EDS1 is indeed necessary for PAD4's function in immunity. However, showing that EDS1 is activated by PAD4 and is sufficient for mediating PAD4's effect on immunity is not easy: we would need a means to specifically activate EDS1. It is not known how to produce a constitutively active version of EDS1 or how to activate EDS1 directly in some other way. We also discussed another possibility, aside from the PAD4 mediator hypothesis: the apparent lack of a *pad4-1* effect on the transcriptome could be flg22-specific. To test this hypothesis, *pad4-1* could be profiled after elf18 treatment, since our previous work (Kim et al. 2014) showed that elf18 stimulates the network in very different ways than flg22.

5.3 Concluding Remarks

Producing and validating mechanistic models of the summarized immune signaling network, the JA-ET-PAD4-SA network, would be a major advance in molecular systems biology—the first of its kind for molecular genetics. Although it appears that several more rounds of experimental data collection and modeling will likely be needed to discover and validate such models, here are some suggestions of what could be done next with a good mechanistic model. A major feature of a good mechanistic model is that it would be able to predict dynamic network behavior given only network inputs. One could then look for network inputs that produce network behaviors of interest. For example, under what input conditions does SA and JA behavior appear synergistic, and under what input conditions do they appear antagonistic? Further, we could ask the model whether high SA and high JA levels can stably coexist in the network. From a summary network motif (Kim et al. 2014), we predicted that such a state might exist. If high SA and high JA can

indeed stably coexist in our mechanistic models, we could look for the set of all network inputs that produce this network state. Discovered or designed network inputs that behave like this family of inputs might confer effective broad spectrum immunity against both biotrophic pathogens (SA-mediated defenses) and necrotrophic pathogens (JA-mediated defenses). The current literature perspective of antagonism between SA and JA may be only one network mode of many.

I found an unexpected delight while in digging in our flg22 transcriptome and hormone data for the work detailed in Chapter 4. The more questions I asked of the data, the more I found answers that were surprising and at times contradictory to current literature interpretations of sector functions. I grew increasingly convinced that a network reconstitution approach produces powerful datasets that can unravel and resolve the mechanisms of even a very complex signaling network, the plant immune signaling network. The empirical approach of defining a complex network in terms of factors that have large effects on network function, and then combinatorially disabling these factors, has the potential to help crack complex problems in systems biology.

Bibliography

Aerts, Stein, Peter Van Loo, Gert Thijs, Yves Moreau, and Bart De Moor. 2003. "Computational Detection of Cis -Regulatory Modules." *Bioinformatics (Oxford, England)* 19 Suppl 2 (October): ii5–14.

Alabadí, D, T Oyama, M J Yanovsky, F G Harmon, P Más, and S A Kay. 2001. "Reciprocal Regulation between TOC1 and LHY/CCA1 within the Arabidopsis Circadian Clock." *Science (New York, N.Y.)* 293 (5531). American Association for the Advancement of Science: 880–83. doi:10.1126/science.1061320.

Alon, Uri. 2006. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. 1st ed. Chapman and Hall/CRC.

Alonso, J M, T Hirayama, G Roman, S Nourizadeh, and J R Ecker. 1999. "EIN2, a Bifunctional Transducer of Ethylene and Stress Responses in Arabidopsis." *Science (New York, N.Y.)* 284 (5423): 2148–52.

"aneurIST." Accessed February 26, 2015. http://www.aneurist.org/.

Asai, Shuta, and Ken Shirasu. 2015. "Plant Cells under Siege: Plant Immune System versus Pathogen Effectors." *Current Opinion in Plant Biology* 28 (September): 1–8. doi:10.1016/j.pbi.2015.08.008.

Ashburner, M, C A Ball, J A Blake, D Botstein, H Butler, J M Cherry, A P Davis, et al. 2000. "Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium." *Nature Genetics* 25 (1). Nature America Inc.: 25–29. doi:10.1038/75556.

Bailey, T L, and M Gribskov. 1998. "Combining Evidence Using P-Values: Application to Sequence Homology Searches." *Bioinformatics (Oxford, England)* 14 (1): 48–54.

Banerjee, Subhasis, and Indrani Bose. 2011. "Transient Pulse Formation in Jasmonate Signaling Pathway." *Journal of Theoretical Biology* 273 (1): 188–96. doi:10.1016/j.jtbi.2010.12.037.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1). doi:10.18637/jss.v067.i01.

Bebber, Daniel P, and Sarah J Gurr. 2015. "Crop-Destroying Fungal and Oomycete Pathogens Challenge Food Security." *Fungal Genetics and Biology : FG & B* 74 (January): 62–64. doi:10.1016/j.fgb.2014.10.012.

Bebber, Daniel P., Mark A. T. Ramotowski, and Sarah J. Gurr. 2013. "Crop Pests and Pathogens Move Polewards in a Warming World." *Nature Climate Change* 3 (11). Nature Publishing Group: 985–88. doi:10.1038/nclimate1990.

Bigeard, Jean, Jean Colcombet, and Heribert Hirt. 2015. "Signaling Mechanisms in Pattern-Triggered Immunity (PTI)." *Molecular Plant* 8 (4): 521–39. doi:10.1016/j.molp.2014.12.022.

"BioModels." Accessed February 26, 2015. http://www.ebi.ac.uk/biomodels-main/modelmonth.

Blankenberg, Daniel, Assaf Gordon, Gregory Von Kuster, Nathan Coraor, James Taylor, and Anton Nekrutenko. 2010. "Manipulation of FASTQ Data with Galaxy." *Bioinformatics (Oxford, England)* 26 (14): 1783–85. doi:10.1093/bioinformatics/btq281.

Blankenberg, Daniel, Gregory Von Kuster, Nathaniel Coraor, Guruprasad Ananda, Ross Lazarus, Mary Mangan, Anton Nekrutenko, and James Taylor. 2010. "Galaxy: A Web-Based Genome Analysis Tool for Experimentalists." *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]* Chapter 19 (January):

Unit 19.10.1–21. doi:10.1002/0471142727.mb1910s89.
Boutrot, Freddy, Cécile Segonzac, Katherine N Chang, Hong Qiao, Joseph R Ecker, Cyril Zipfel, and John P Rathjen. 2010. “Direct Transcriptional Control of the Arabidopsis Immune Receptor FLS2 by the Ethylene-Dependent Transcription Factors EIN3 and EIL1.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (32): 14502–7. doi:10.1073/pnas.1003347107.
Broekgaarden, Colette, Lotte Caarls, Irene A Vos, Corné M J Pieterse, and Saskia C M Van Wees. 2015. “Ethylene: Traffic Controller on Hormonal Crossroads to Defense.” *Plant Physiology* [Epub ahea (October). doi:10.1104/pp.15.01020.
Burnham, K. P. 2004. “Multimodel Inference: Understanding AIC and BIC in Model Selection.” *Sociological Methods & Research* 33 (2): 261–304. doi:10.1177/0049124104268644.
Chang, Katherine Noelani, Shan Zhong, Matthew T Weirauch, Gary Hon, Mattia Pelizzola, Hai Li, Shao-Shan Carol Huang, et al. 2013. “Temporal Transcriptional Response to Ethylene Gas Drives Growth Hormone Cross-Regulation in Arabidopsis.” *eLife* 2 (January): e00675. doi:10.7554/eLife.00675.
Chen, Huamin, Li Xue, Satya Chintamanani, Hugo Germain, Huiqiong Lin, Haitao Cui, Run Cai, et al. 2009. “ETHYLENE INSENSITIVE3 and ETHYLENE INSENSITIVE3-LIKE1 Repress SALICYLIC ACID INDUCTION DEFICIENT2 Expression to Negatively Regulate Plant Innate Immunity in Arabidopsis.” *The Plant Cell* 21 (8): 2527–40. doi:10.1105/tpc.108.065193.
Chickarmane, Vijay S, Sean P Gordon, Paul T Tarr, Marcus G Heisler, and Elliot M Meyerowitz. 2012. “Cytokinin Signaling as a Positional Cue for Patterning the Apical-Basal Axis of the Growing Arabidopsis Shoot Meristem.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (10): 4002–7. doi:10.1073/pnas.1200636109.
Choi, Kyuha, Xiaohui Zhao, Krystyna A Kelly, Oliver Venn, James D Higgins, Nataliya E Yelina, Thomas J Hardcastle, et al. 2013. “Arabidopsis Meiotic Crossover Hot Spots Overlap with H2A.Z Nucleosomes at Gene Promoters.” *Nature Genetics* 45 (11): 1327–36. doi:10.1038/ng.2766.
Coe, Robert. 2002. “It’s the Effect Size, Stupid: What Effect Size Is and Why It Is Important,” (paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England, 12-14 September, 2002). http://www.leeds.ac.uk/educol/documents/00002182.htm.
Costanzo, Michael, Anastasia Baryshnikova, Jeremy Bellay, Yungil Kim, Eric D Spear, Carolyn S Sevier, Huiming Ding, et al. 2010. “The Genetic Landscape of a Cell.” *Science (New York, N.Y.)* 327 (5964): 425–31. doi:10.1126/science.1180823.
Dalchau, N., K. E. Hubbard, F. C. Robertson, C. T. Hotta, H. M. Briggs, G.-B. Stan, J. M. Goncalves, and A. A. R. Webb. 2010. “Correct Biological Timing in Arabidopsis Requires Multiple Light-Signaling Pathways.” *Proceedings of the National Academy of Sciences* 107 (29): 13171–76. doi:10.1073/pnas.1001429107.
Dangl, J L, and J D Jones. 2001. “Plant Pathogens and Integrated Defence Responses to Infection.” *Nature* 411 (6839): 826–33. doi:10.1038/35081161.
Davuluri, Ramana V, Hao Sun, Saranyan K Palaniswamy, Nicole Matthews, Carlos Molina, Mike Kurtz, and Erich Grotewold. 2003. “AGRIS: Arabidopsis Gene Regulatory Information Server, an Information Resource of Arabidopsis Cis-Regulatory Elements and Transcription Factors.” *BMC Bioinformatics* 4 (June): 25. doi:10.1186/1471-2105-4-25.
Deng, Weiwei, Diana M Buzas, Hua Ying, Masumi Robertson, Jennifer Taylor, William

James Peacock, Elizabeth S Dennis, and Chris Helliwell. 2013. “Arabidopsis Polycomb Repressive Complex 2 Binding Sites Contain Putative GAGA Factor Binding Motifs within Coding Regions of Genes.” *BMC Genomics* 14 (1). BioMed Central Ltd: 593. doi:10.1186/1471-2164-14-593.

Dillies, Marie-Agnès, Andrea Rau, Julie Aubert, Christelle Hennequet-Antier, Marine Jeanmougin, Nicolas Servant, Céline Keime, et al. 2013. “A Comprehensive Evaluation of Normalization Methods for Illumina High-Throughput RNA Sequencing Data Analysis.” *Briefings in Bioinformatics* 14 (6): 671–83. doi:10.1093/bib/bbs046.

Ding, Bo, and Guo-Liang Wang. 2015. “Chromatin versus Pathogens: The Function of Epigenetics in Plant Immunity.” *Frontiers in Plant Science* 6 (January): 675. doi:10.3389/fpls.2015.00675.

Dodd, Antony N, Neil Dalchau, Michael J Gardner, Seong-Jin Baek, and Alex A R Webb. 2014. “The Circadian Clock Has Transient Plasticity of Period and Is Required for Timing of Nocturnal Processes in Arabidopsis.” *The New Phytologist* 201 (1): 168–79. doi:10.1111/nph.12489.

Dodds, Peter N, and John P Rathjen. 2010. “Plant Immunity: Towards an Integrated View of Plant-Pathogen Interactions.” *Nature Reviews. Genetics* 11 (8): 539–48. doi:10.1038/nrg2812.

Economou, Andrew D, Atsushi Ohazama, Thantrira Porntaveetus, Paul T Sharpe, Shigeru Kondo, M Albert Basson, Amel Gritli-Linde, Martyn T Cobourne, and Jeremy B A Green. 2012. “Periodic Stripe Formation by a Turing Mechanism Operating at Growth Zones in the Mammalian Palate.” *Nature Genetics* 44 (3): 348–51. doi:10.1038/ng.1090.

Efron, Bradley, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. 2015. “Least Angle Regression.” *The Annals of Statistics* 32 (2). Institute of Mathematical Statistics: 407–99.

Eichten, Steven R, Roman Briskine, Jawon Song, Qing Li, Ruth Swanson-Wagner, Peter J Hermanson, Amanda J Waters, et al. 2013. “Epigenetic and Genetic Influences on DNA Methylation Variation in Maize Populations.” *The Plant Cell* 25 (8): 2783–97. doi:10.1105/tpc.113.114793.

Ellner, Stephen P., and John Guckenheimer. 2011. *Dynamic Models in Biology*. Princeton University Press.

Feys, B J, L J Moisan, M A Newman, and J E Parker. 2001. “Direct Interaction between the Arabidopsis Disease Resistance Signaling Proteins, EDS1 and PAD4.” *The EMBO Journal* 20 (19): 5400–5411. doi:10.1093/emboj/20.19.5400.

Fischer, Hans Peter. 2008. “Mathematical Modeling of Complex Biological Systems: From Parts Lists to Understanding Systems Behavior.” *Alcohol Research & Health : The Journal of the National Institute on Alcohol Abuse and Alcoholism* 31 (1): 49–59.

Freedman, D.H. 2004. “The Virtual Heart.” *Technology Review*.

Garcion, Christophe, Antje Lohmann, Elisabeth Lamodière, Jérémy Catinot, Antony Buchala, Peter Doermann, and Jean-Pierre Métraux. 2008. “Characterization and Biological Function of the ISOCHORISMATE SYNTHASE2 Gene of Arabidopsis.” *Plant Physiology* 147 (3): 1279–87. doi:10.1104/pp.108.119420.

Gendron, Joshua M, José L Pruneda-Paz, Colleen J Doherty, Andrew M Gross, S Earl Kang, and Steve A Kay. 2012. “Arabidopsis Circadian Clock Protein, TOC1, Is a DNA-Binding Transcription Factor.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (8): 3167–72.

doi:10.1073/pnas.1200355109.
Giardine, Belinda, Cathy Riemer, Ross C Hardison, Richard Burhans, Laura Elnitski, Prachi Shah, Yi Zhang, et al. 2005. “Galaxy: A Platform for Interactive Large-Scale Genome Analysis.” *Genome Research* 15 (10): 1451–55. doi:10.1101/gr.4086505.
Glazebrook, J, M Zook, F Mert, I Kagan, E E Rogers, I R Crute, E B Holub, R Hammerschmidt, and F M Ausubel. 1997. “Phytoalexin-Deficient Mutants of Arabidopsis Reveal That PAD4 Encodes a Regulatory Factor and That Four PAD Genes Contribute to Downy Mildew Resistance.” *Genetics* 146 (1): 381–92.
Goecks, Jeremy, Anton Nekrutenko, and James Taylor. 2010. “Galaxy: A Comprehensive Approach for Supporting Accessible, Reproducible, and Transparent Computational Research in the Life Sciences.” *Genome Biology* 11 (8): R86. doi:10.1186/gb-2010-11-8-r86.
Gordon, Sean P, Vijay S Chickarmane, Carolyn Ohno, and Elliot M Meyerowitz. 2009. “Multiple Feedback Loops through Cytokinin Signaling Control Stem Cell Number within the Arabidopsis Shoot Meristem.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (38): 16529–34. doi:10.1073/pnas.0908122106.
Guo, Hongwei, and Joseph R Ecker. 2003. “Plant Responses to Ethylene Gas Are Mediated by SCFEBF1/EBF2-Dependent Proteolysis of EIN3 Transcription Factor.” *Cell* 115 (6): 667–77. doi:10.1016/S0092-8674(03)00969-3.
Hartman, J L 4th, B Garvik, and L Hartwell. 2001. “Principles for the Buffering of Genetic Variation.” *Science (New York, N.Y.)* 291 (5506): 1001–4.
Hecker, Andreas, Luise H Brand, Sébastien Peter, Nathalie Simoncello, Joachim Kilian, Klaus Harter, Valérie Gaudin, and Dierk Wanke. 2015. “The Arabidopsis GAGA-Binding Factor BASIC PENTACYSTEINE6 Recruits the POLYCOMB-REPRESSIVE COMPLEX1 Component LIKE HETEROCHROMATIN PROTEIN1 to GAGA DNA Motifs.” *Plant Physiology* 168 (3): 1013–24. doi:10.1104/pp.15.00409.
Hillmer, Rachel A. 2015. “Systems Biology for Biologists.” *PLoS Pathogens* 11 (5). Public Library of Science: e1004786. doi:10.1371/journal.ppat.1004786.
Hillmer, Rachel A., and Fumiaki Katagiri. 2016. “Toward Predictive Modeling of Large and Complex Biological Signaling Networks.” *Physiological and Molecular Plant Pathology*, January. doi:10.1016/j.pmpp.2016.01.009.
Hurvich, Clifford M., and Chih-Ling Tsai. 1989. “Regression and Time Series Model Selection in Small Samples.” *Biometrika* 76 (2): 297–307. doi:10.1093/biomet/76.2.297.
Jaeger, Katja E, Nick Pullen, Sergey Lamzin, Richard J Morris, and Philip A Wigge. 2013. “Interlocking Feedback Loops Govern the Dynamic Behavior of the Floral Transition in Arabidopsis.” *The Plant Cell* 25 (3): 820–33. doi:10.1105/tpc.113.109355.
Jansen, Ritsert C. 2003. “Studying Complex Biological Systems Using Multifactorial Perturbation.” *Nature Reviews Genetics* 4 (2): 145–51. doi:10.1038/nrg996.
Jirage, D, T L Tootle, T L Reuber, L N Frost, B J Feys, J E Parker, F M Ausubel, and J Glazebrook. 1999. “Arabidopsis Thaliana PAD4 Encodes a Lipase-like Gene That Is Important for Salicylic Acid Signaling.” *Proceedings of the National Academy of Sciences of the United States of America* 96 (23): 13583–88.
Jones, Jonathan D G, and Jeffery L Dangl. 2006. “The Plant Immune System.” *Nature* 444 (7117): 323–29. doi:10.1038/nature05286.
Keeling, Matt J. 2005. “Models of Foot-and-Mouth Disease.” *Proceedings. Biological Sciences / The Royal Society* 272 (1569): 1195–1202. doi:10.1098/rspb.2004.3046.

Kim, Yungil, Kenichi Tsuda, Daisuke Igarashi, Rachel A Hillmer, Hitoshi Sakakibara, Chad L Myers, and Fumiaki Katagiri. 2014. “Mechanisms Underlying Robustness and Tunability in a Plant Immune Signaling Network.” *Cell Host & Microbe* 15 (1): 84–94. doi:10.1016/j.chom.2013.12.002.

Kitano, Hiroaki. 2002. “Systems Biology: A Brief Overview.” *Science (New York, N.Y.)* 295 (5560): 1662–64. doi:10.1126/science.1069492.

Kohl, Peter, and Denis Noble. 2009. “Systems Biology and the Virtual Physiological Human.” *Molecular Systems Biology* 5 (January): 292. doi:10.1038/msb.2009.51.

Kojima, Mikiko, Tomoe Kamada-Nobusada, Hirokazu Komatsu, Kentaro Takei, Takeshi Kuroha, Masaharu Mizutani, Motoyuki Ashikari, et al. 2009. “Highly Sensitive and High-Throughput Analysis of Plant Hormones Using MS-Probe Modification and Liquid Chromatography-Tandem Mass Spectrometry: An Application for Hormone Profiling in Oryza Sativa.” *Plant & Cell Physiology* 50 (7): 1201–14. doi:10.1093/pcp/pcp057.

Kojima, Mikiko, and Hitoshi Sakakibara. 2012. “Highly Sensitive High-Throughput Profiling of Six Phytohormones Using MS-Probe Modification and Liquid Chromatography-Tandem Mass Spectrometry.” *Methods in Molecular Biology (Clifton, N.J.)* 918 (January): 151–64. doi:10.1007/978-1-61779-995-2_11.

Kosmrlj, Andrej, Elizabeth L Read, Ying Qi, Todd M Allen, Marcus Altfeld, Steven G Deeks, Florencia Pereyra, Mary Carrington, Bruce D Walker, and Arup K Chakraborty. 2010. “Effects of Thymic Selection of the T-Cell Repertoire on HLA Class I-Associated Control of HIV Infection.” *Nature* 465 (7296): 350–54. doi:10.1038/nature08997.

Krishnakumar, Vivek, Matthew R Hanlon, Sergio Contrino, Erik S Ferlanti, Svetlana Karamycheva, Maria Kim, Benjamin D Rosen, et al. 2015. “Araport: The Arabidopsis Information Portal.” *Nucleic Acids Research* 43 (Database issue): D1003–9. doi:10.1093/nar/gku1200.

Laisk, Agu, Hillar Eichelmann, and Vello Oja. 2006. “C3 Photosynthesis in Silico.” *Photosynthesis Research* 90 (1): 45–66. doi:10.1007/s11120-006-9109-1.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L Salzberg. 2009. “Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome.” *Genome Biology* 10 (3): R25. doi:10.1186/gb-2009-10-3-r25.

Le Novère, N. “Modelling Success Stories in Systems Biology.” Accessed February 25, 2015. https://nlenov.wordpress.com/2013/02/19/modelling-success-stories-in-systems-biology/.

Libourel, Igor G L, and Yair Shachar-Hill. 2008. “Metabolic Flux Analysis in Plants: From Intelligent Design to Rational Engineering.” *Annual Review of Plant Biology* 59 (January). Annual Reviews: 625–50. doi:10.1146/annurev.arplant.58.032806.103822.

Lim, Lee P, Nelson C Lau, Philip Garrett-Engele, Andrew Grimson, Janell M Schelter, John Castle, David P Bartel, Peter S Linsley, and Jason M Johnson. 2005. “Microarray Analysis Shows That Some microRNAs Downregulate Large Numbers of Target mRNAs.” *Nature* 433 (7027): 769–73. doi:10.1038/nature03315.

Locke, J C W, A J Millar, and M S Turner. 2005. “Modelling Genetic Networks with Noisy and Varied Experimental Data: The Circadian Clock in Arabidopsis Thaliana.” *Journal of Theoretical Biology* 234 (3): 383–93. doi:10.1016/j.jtbi.2004.11.038.

Locke, James C W, László Kozma-Bognár, Peter D Gould, Balázs Fehér, Eva Kevei, Ferenc Nagy, Matthew S Turner, Anthony Hall, and Andrew J Millar. 2006. “Experimental Validation of a Predicted Feedback Loop in the Multi-Oscillator Clock

of Arabidopsis Thaliana." *Molecular Systems Biology* 2 (January): 59. doi:10.1038/msb4100102.
Locke, James C W, Megan M Southern, László Kozma-Bognár, Victoria Hibberd, Paul E Brown, Matthew S Turner, and Andrew J Millar. 2005. "Extension of a Genetic Network Model by Iterative Experimentation and Mathematical Analysis." *Molecular Systems Biology* 1 (January): 2005.0013. doi:10.1038/msb4100018.
Lorenzo, Oscar, Raquel Piqueras, Jose J Sánchez-Serrano, and Roberto Solano. 2003. "ETHYLENE RESPONSE FACTOR1 Integrates Signals from Ethylene and Jasmonate Pathways in Plant Defense." *The Plant Cell* 15 (1): 165–78.
Maarleveld, Timo R, Ruchir A Khandelwal, Brett G Olivier, Bas Teusink, and Frank J Bruggeman. 2013. "Basic Concepts and Principles of Stoichiometric Modeling of Metabolic Networks." *Biotechnology Journal* 8 (9): 997–1008. doi:10.1002/biot.201200291.
Markakis, Marios Nektarios, Tinne De Cnodder, Michal Lewandowski, Damien Simon, Agnieszka Boron, Daria Balcerowicz, Thanaa Doubbo, et al. 2012. "Identification of Genes Involved in the ACC-Mediated Control of Root Cell Elongation in Arabidopsis Thaliana." *BMC Plant Biology* 12 (January): 208. doi:10.1186/1471-2229-12-208.
Mateos, Julieta L, Pedro Madrigal, Kenichi Tsuda, Vimal Rawat, René Richter, Maida Romera-Branchat, Fabio Fornara, Korbinian Schneeberger, Paweł Krajewski, and George Coupland. 2015. "Combinatorial Activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C Define Distinct Modes of Flowering Regulation in Arabidopsis." *Genome Biology* 16 (1). BioMed Central Ltd: 31. doi:10.1186/s13059-015-0597-1.
McNeil, S. D. 2000. "Metabolic Modeling Identifies Key Constraints on an Engineered Glycine Betaine Synthesis Pathway in Tobacco." *PLANT PHYSIOLOGY* 124 (1): 153–62. doi:10.1104/pp.124.1.153.
Molina, Carlos, and Erich Grotewold. 2005. "Genome Wide Analysis of Arabidopsis Core Promoters." *BMC Genomics* 6 (1). BioMed Central Ltd: 25. doi:10.1186/1471-2164-6-25.
Mur, Luis A J, Paul Kenton, Rainer Atzorn, Otto Miersch, and Claus Wasternack. 2006. "The Outcomes of Concentration-Specific Interactions between Salicylate and Jasmonate Signaling Include Synergy, Antagonism, and Oxidative Stress Leading to Cell Death." *Plant Physiology* 140 (1): 249–62. doi:10.1104/pp.105.072348.
Nägele, Thomas, Sebastian Henkel, Imke Hörmiller, Thomas Sauter, Oliver Sawodny, Michael Ederer, and Arnd G Heyer. 2010. "Mathematical Modeling of the Central Carbohydrate Metabolism in Arabidopsis Reveals a Substantial Regulatory Influence of Vacuolar Invertase on Whole Plant Carbon Metabolism." *Plant Physiology* 153 (1): 260–72. doi:10.1104/pp.110.154443.
Nawrot, Robert, Jakub Barylski, Grzegorz Nowicki, Justyna Broniarczyk, Waldemar Buchwald, and Anna Goździcka-Józefiak. 2014. "Plant Antimicrobial Peptides." *Folia Microbiologica* 59 (3): 181–96. doi:10.1007/s12223-013-0280-4.
Newman, Mari-Anne, Thomas Sundelin, Jon T. Nielsen, and Gitte Erbs. 2013. "MAMP (microbe-Associated Molecular Pattern) Triggered Immunity in Plants." *Frontiers in Plant Science* 4 (January): 139. doi:10.3389/fpls.2013.00139.
Oerke, E.-C. 2005. "Crop Losses to Pests." *The Journal of Agricultural Science* 144 (01). Cambridge University Press: 31. doi:10.1017/S0021859605005708.
Ouellette, J. 2013. "Biologists Home in on Turing Patterns." *Quanta Magazine*. https://www.quantamagazine.org/20130325-biologists-home-in-on-turing-patterns/.
Pajerowska-Mukhtar, Karolina M, Wei Wang, Yasuomi Tada, Nodoka Oka, Chandra L

Tucker, Jose Pedro Fonseca, and Xinnian Dong. 2012. “The HSF-like Transcription Factor TBF1 Is a Major Molecular Switch for Plant Growth-to-Defense Transition.” *Current Biology : CB* 22 (2): 103–12. doi:10.1016/j.cub.2011.12.015.

Park, Joon-Hyun, Rayko Halitschke, Ho Bang Kim, Ian T Baldwin, Kenneth A Feldmann, and René Feyereisen. 2002. “A Knock-out Mutation in Allene Oxide Synthase Results in Male Sterility and Defective Wound Signal Transduction in Arabidopsis due to a Block in Jasmonic Acid Biosynthesis.” *The Plant Journal : For Cell and Molecular Biology* 31 (1): 1–12.

Pegadaraju, Venkatramana, Caleb Knepper, John Reese, and Jyoti Shah. 2005. “Premature Leaf Senescence Modulated by the Arabidopsis PHYTOALEXIN DEFICIENT4 Gene Is Associated with Defense against the Phloem-Feeding Green Peach Aphid.” *Plant Physiology* 139 (4): 1927–34. doi:10.1104/pp.105.070433.

Penninckx, I A, B P Thomma, A Buchala, J P Métraux, and W F Broekaert. 1998. “Concomitant Activation of Jasmonate and Ethylene Response Pathways Is Required for Induction of a Plant Defensin Gene in Arabidopsis.” *The Plant Cell* 10 (12): 2103–13.

Pettersson, Gosta, and Ulf Ryde-Pettersson. 1988. “A Mathematical Model of the Calvin Photosynthesis Cycle.” *European Journal of Biochemistry* 175 (3): 661–72. doi:10.1111/j.1432-1033.1988.tb14242.x.

Pieterse, Corné M J, Antonio Leon-Reyes, Sjoerd Van der Ent, and Saskia C M Van Wees. 2009. “Networking by Small-Molecule Hormones in Plant Immunity.” *Nature Chemical Biology* 5 (5): 308–16. doi:10.1038/nchembio.164.

Pieterse, Corné M J, Dieuwertje Van der Does, Christos Zamioudis, Antonio Leon-Reyes, and Saskia C M Van Wees. 2012. “Hormonal Modulation of Plant Immunity.” *Annual Review of Cell and Developmental Biology* 28 (January): 489–521. doi:10.1146/annurev-cellbio-092910-154055.

Pigozzo, Alexandre Bittencourt, Gilson Costa Macedo, Rodrigo Weber dos Santos, and Marcelo Lobosco. 2013. “On the Computational Modeling of the Innate Immune System.” *BMC Bioinformatics* 14 Suppl 6 (6). BioMed Central: S7. doi:10.1186/1471-2105-14-S6-S7.

Poolman, M. G. 2000. “Modelling Photosynthesis and Its Control.” *Journal of Experimental Botany* 51 (90001): 319–28. doi:10.1093/jexbot/51.suppl_1.319.

Pré, Martial, Mirna Atallah, Antony Champion, Martin De Vos, Corné M J Pieterse, and Johan Memelink. 2008. “The AP2/ERF Domain Transcription Factor ORA59 Integrates Jasmonic Acid and Ethylene Signals in Plant Defense.” *Plant Physiology* 147 (3): 1347–57. doi:10.1104/pp.108.117523.

Rai, Muneeza Iqbal, Xiaomin Wang, Derek M Thibault, Hyo Jung Kim, Matthew M Bombyk, Brad M Binder, Samina N Shakeel, and G Eric Schaller. 2015. “The ARGOS Gene Family Functions in a Negative Feedback Loop to Desensitize Plants to Ethylene.” *BMC Plant Biology* 15 (January): 157. doi:10.1186/s12870-015-0554-x.

Rallapalli, Ghanasyam, Eric M Kemen, Alexandre Robert-Seilaniantz, Cécile Segonzac, Graham J Etherington, Kee Hoon Sohn, Daniel MacLean, and Jonathan D G Jones. 2014. “EXPRSS: An Illumina Based High-Throughput Expression-Profiling Method to Reveal Transcriptional Dynamics.” *BMC Genomics* 15 (January): 341. doi:10.1186/1471-2164-15-341.

Rapaport, Franck, Raya Khanin, Yupu Liang, Mono Pirun, Azra Krek, Paul Zumbo, Christopher E Mason, Nicholas D Socci, and Doron Betel. 2013. “Comprehensive Evaluation of Differential Gene Expression Analysis Methods for RNA-Seq Data.”

Genome Biology 14 (9): R95. doi:10.1186/gb-2013-14-9-r95.
Reuber, T L, J M Plotnikova, J Dewdney, E E Rogers, W Wood, and F M Ausubel. 1998. "Correlation of Defense Gene Induction Defects with Powdery Mildew Susceptibility in Arabidopsis Enhanced Disease Susceptibility Mutants." *The Plant Journal : For Cell and Molecular Biology* 16 (4): 473–85.
Robatzek, Silke, Delphine Chinchilla, and Thomas Boller. 2006. "Ligand-Induced Endocytosis of the Pattern Recognition Receptor FLS2 in Arabidopsis." *Genes & Development* 20 (5): 537–42. doi:10.1101/gad.366506.
Rutherford, S L. 2000. "From Genotype to Phenotype: Buffering Mechanisms and the Storage of Genetic Information." *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* 22 (12): 1095–1105. doi:10.1002/1521-1878(200012)22:12<1095::AID-BIES7>3.0.CO;2-A.
Shampine, L.F., and S. Thompson. 2001. "Solving DDEs in Matlab." *Applied Numerical Mathematics* 37 (4): 441–58. doi:10.1016/S0168-9274(00)00055-6.
Sinden, Richard R, Vladimir N Potaman, Elena A Oussatcheva, Christopher E Pearson, Yuri L Lyubchenko, and Luda S Shlyakhtenko. 2002. "Triplet Repeat DNA Structures and Human Genetic Disease: Dynamic Mutations from Dynamic DNA." *Journal of Biosciences* 27 (1 Suppl 1): 53–65.
Storey, John D. 2015. "Qvalue: Q-Value Estimation for False Discovery Rate Control."
Storey, John D, and Robert Tibshirani. 2003. "Statistical Significance for Genomewide Studies." *Proceedings of the National Academy of Sciences of the United States of America* 100 (16): 9440–45. doi:10.1073/pnas.1530509100.
Strange, Richard N, and Peter R Scott. 2005. "Plant Disease: A Threat to Global Food Security." *Annual Review of Phytopathology* 43 (January). Annual Reviews: 83–116. doi:10.1146/annurev.phyto.43.113004.133839.
Székely, Tamás, and Kevin Burrage. 2014. "Stochastic Simulation in Systems Biology." *Computational and Structural Biotechnology Journal* 12 (20-21): 14–25. doi:10.1016/j.csbj.2014.10.003.
The Gene Ontology Consortium. 2014. "Gene Ontology Consortium: Going Forward." *Nucleic Acids Research* 43 (D1): D1049–56. doi:10.1093/nar/gku1179.
Tibshirani, Robert. 2011. "Regression Shrinkage and Selection via the Lasso: A Retrospective." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (3): 273–82. doi:10.1111/j.1467-9868.2011.00771.x.
Truman, William, and Jane Glazebrook. 2012. "Co-Expression Analysis Identifies Putative Targets for CBP60g and SARD1 Regulation." *BMC Plant Biology* 12 (January): 216. doi:10.1186/1471-2229-12-216.
Tsuda, Kenichi, and Fumiaki Katagiri. 2010. "Comparing Signaling Mechanisms Engaged in Pattern-Triggered and Effector-Triggered Immunity." *Current Opinion in Plant Biology* 13 (4): 459–65. doi:10.1016/j.pbi.2010.04.006.
Tsuda, Kenichi, Masanao Sato, Jane Glazebrook, Jerry D Cohen, and Fumiaki Katagiri. 2008. "Interplay between MAMP-Triggered and SA-Mediated Defense Responses." *The Plant Journal : For Cell and Molecular Biology* 53 (5): 763–75. doi:10.1111/j.1365-313X.2007.03369.x.
Tsuda, Kenichi, Masanao Sato, Thomas Stoddard, Jane Glazebrook, and Fumiaki Katagiri. 2009. "Network Properties of Robust Immunity in Plants." *PLoS Genetics* 5 (12): e1000772. doi:10.1371/journal.pgen.1000772.
Turing, A. M. 1952. "The Chemical Basis of Morphogenesis." *Philosophical Transactions of the Royal Society B: Biological Sciences* 237 (641): 37–72. doi:10.1098/rstb.1952.0012.

Underwood, William. 2012. “The Plant Cell Wall: A Dynamic Barrier against Pathogen Invasion.” *Frontiers in Plant Science* 3 (January): 85. doi:10.3389/fpls.2012.00085.
Van der Does, Dieuwertje, Antonio Leon-Reyes, Annemart Koornneef, Marcel C Van Verk, Nicole Rodenburg, Laurens Pauwels, Alain Goossens, et al. 2013. “Salicylic Acid Suppresses Jasmonic Acid Signaling Downstream of SCFCOI1-JAZ by Targeting GCC Promoter Motifs via Transcription Factor ORA59.” *The Plant Cell* 25 (2): 744–61. doi:10.1105/tpc.112.108548.
Verhage, Adriaan, Ido Vlaardingerbroek, Ciska Raaymakers, Nicole M Van Dam, Marcel Dicke, Saskia C M Van Wees, and Corné M J Pieterse. 2011. “Rewiring of the Jasmonate Signaling Pathway in Arabidopsis during Insect Herbivory.” *Frontiers in Plant Science* 2 (January): 47. doi:10.3389/fpls.2011.00047.
von Dassow, George, Eli Meir, Edwin M Munro, and Garrett M Odell. 2000. “The Segment Polarity Network Is a Robust Developmental Module.” *Nature* 406 (6792). Macmillian Magazines Ltd.: 188–92. doi:10.1038/35018085.
von Dassow, George, and Garrett M Odell. 2002. “Design and Constraints of the Drosophila Segment Polarity Module: Robust Spatial Patterning Emerges from Intertwined Cell State Switches.” *The Journal of Experimental Zoology* 294 (3): 179–215. doi:10.1002/jez.10144.
von Malek, Bernadette, Eric van der Graaff, Kay Schneitz, and Beat Keller. 2002. “The Arabidopsis Male-Sterile Mutant dde2-2 Is Defective in the ALLENE OXIDE SYNTHASE Gene Encoding One of the Key Enzymes of the Jasmonic Acid Biosynthesis Pathway.” *Planta* 216 (1): 187–92. doi:10.1007/s00425-002-0906-2.
Vos, Irene A., Corné M.J. Pieterse, and Saskia C.M. Van Wees. 2013. “Costs and Benefits of Hormone-Regulated Plant Defences. Plant Pathology.” *Plant Pathology*, June.
Wang, Lin, Raka M Mitra, Keegan D Hasselmann, Masanao Sato, Lisa Lenarz-Wyatt, Jerry D Cohen, Fumiaki Katagiri, and Jane Glazebrook. 2008. “The Genetic Network Controlling the Arabidopsis Transcriptional Response to Pseudomonas Syringae Pv. Maculicola: Roles of Major Regulators and the Phytotoxin Coronatine.” *Molecular Plant-Microbe Interactions : MPMI* 21 (11): 1408–20. doi:10.1094/MPMI-21-11-1408.
Wang, Lin, Kenichi Tsuda, William Truman, Masanao Sato, Le V Nguyen, Fumiaki Katagiri, and Jane Glazebrook. 2011. “CBP60g and SARD1 Play Partially Redundant Critical Roles in Salicylic Acid Signaling.” *The Plant Journal : For Cell and Molecular Biology* 67 (6): 1029–41. doi:10.1111/j.1365-313X.2011.04655.x.
Wang, Shui, Wendy E Durrant, Junqi Song, Natalie W Spivey, and Xinnian Dong. 2010. “Arabidopsis BRCA2 and RAD51 Proteins Are Specifically Involved in Defense Gene Transcription during Plant Immune Responses.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (52): 22716–21. doi:10.1073/pnas.1005978107.
Weirauch, Matthew T., Ally Yang, Mihai Albu, Atina G. Cote, Alejandro Montenegro-Montero, Philipp Drewe, Hamed S. Najafabadi, et al. 2014. “Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity.” *Cell* 158 (6): 1431–43. doi:10.1016/j.cell.2014.08.009.
Wildermuth, M C, J Dewdney, G Wu, and F M Ausubel. 2001. “Isochorismate Synthase Is Required to Synthesize Salicylic Acid for Plant Defence.” *Nature* 414 (6863): 562–65. doi:10.1038/35107108.
Winter, Cara M, Ryan S Austin, Servane Blanvillain-Baufumé, Maxwell A Reback, Marie Monniaux, Miin-Feng Wu, Yi Sang, et al. 2011. “LEAFY Target Genes Reveal Floral

Regulatory Logic, Cis Motifs, and a Link to Biotic Stimulus Response." *Developmental Cell* 20 (4): 430–43. doi:10.1016/j.devcel.2011.03.019.
Winter, Debbie, Ben Vinegar, Hardeep Nahal, Ron Ammar, Greg V Wilson, and Nicholas J Provart. 2007. "An 'Electronic Fluorescent Pictograph' Browser for Exploring and Analyzing Large-Scale Biological Data Sets." *PloS One* 2 (8): e718. doi:10.1371/journal.pone.0000718.
Yan, Shunping, Wei Wang, Jorge Marqués, Rajinikanth Mohan, Abdelaty Saleh, Wendy E Durrant, Junqi Song, and Xinnian Dong. 2013. "Salicylic Acid Activates DNA Damage Responses to Potentiate Plant Immunity." *Molecular Cell* 52 (4): 602–10. doi:10.1016/j.molcel.2013.09.019.
Yilmaz, Alper, Maria Katherine Mejia-Guerra, Kyle Kurz, Xiaoyu Liang, Lonnie Welch, and Erich Grotewold. 2011. "AGRIS: The Arabidopsis Gene Regulatory Information Server, an Update." *Nucleic Acids Research* 39 (Database issue): D1118–22. doi:10.1093/nar/gkq1120.
Youssef, Reham M, Margaret H MacDonald, Eric P Brewer, Gary R Bauchan, Kyung-Hwan Kim, and Benjamin F Matthews. 2013. "Ectopic Expression of AtPAD4 Broadens Resistance of Soybean to Soybean Cyst and Root-Knot Nematodes." *BMC Plant Biology* 13 (January): 67. doi:10.1186/1471-2229-13-67.
Zheng, Xiao-Yu, Natalie Weaver Spivey, Weiqing Zeng, Po-Pu Liu, Zheng Qing Fu, Daniel F Klessig, Sheng Yang He, and Xinnian Dong. 2012. "Coronatine Promotes Pseudomonas Syringae Virulence in Plants by Activating a Signaling Cascade That Inhibits Salicylic Acid Accumulation." *Cell Host & Microbe* 11 (6): 587–96. doi:10.1016/j.chom.2012.04.014.
Zhou, N, T L Tootle, F Tsui, D F Klessig, and J Glazebrook. 1998. "PAD4 Functions Upstream from Salicylic Acid to Control Defense Responses in Arabidopsis." *The Plant Cell* 10 (6): 1021–30.
Zhu, Ziqiang, Fengying An, Ying Feng, Pengpeng Li, Li Xue, Mu A, Zhiqiang Jiang, et al. 2011. "Derepression of Ethylene-Stabilized Transcription Factors (EIN3/EIL1) Mediates Jasmonate and Ethylene Signaling Synergy in Arabidopsis." *Proceedings of the National Academy of Sciences of the United States of America* 108 (30): 12539–44. doi:10.1073/pnas.1103959108.
Zipfel, Cyril, Silke Robatzek, Lionel Navarro, Edward J. Oakeley, Jonathan D. G. Jones, Georg Felix, and Thomas Boller. 2004. "Bacterial Disease Resistance in Arabidopsis through Flagellin Perception." *Nature* 428 (6984): 764–67. doi:10.1038/nature02485.